MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 9: 469–499 (2025)
DOI: 10.3897/mbmg.9.165436

Research Article

Surface swabs outperform most traditional honeybee (*Apis mellifera*) hive samples for recovery of eDNA and eRNA

Erin Hill[1], Liz Milla[2], Francisco Encinas-Viso[2], James O'Dwyer[3], Ben Gooden[1], Mariana Hopper[4], John Roberts[1]

1 *CSIRO Health and Biosecurity, Clunies Ross St, Acton, ACT, Australia*
2 *CSIRO Centre for Australian National Biodiversity Research, Clunies Ross St, Acton, ACT, Australia*
3 *CSIRO Australian Centre for Disease Preparedness, Portarlington Rd, East Geelong, VIC, Australia*
4 *CSIRO Health and Biosecurity, Underwood Ave, Floreat, WA, Australia*

Corresponding author: Erin Hill (erinhill19@gmail.com)

Abstract

Genetic sampling from European honeybee (*Apis mellifera*) hives is a promising repository of terrestrial environmental nucleic acids (eNA) from a range of taxonomic sources. eNA samples from honeybee hives can be used to monitor environments via vegetation assessments and plant pathogen detection, while simultaneously assessing honeybee hive health through the identification of honeybee pests and pathogens. However, there is still limited comparison across the different types of hive samples commonly used for biomonitoring. There is also a need for less invasive sampling methods to facilitate adoption of hive eNA surveillance at scale. The aim of this study was to compare the effectiveness of using non-invasive swabbing of hive entrances with common sample types collected from honeybee hives (honey, pollen, and bees) for eNA detection of various taxa. Using a DNA metabarcoding approach for bacteria (16S), animal (COI), fungi (ITS1), and plant (ITS2) identification, and metatranscriptomic sequencing for virus detection, swabbing was found to provide comparable or better detection of taxa compared with more invasive sampling methods. Swabs were most informative for identifying bacterial and fungal communities, while plant detections were similar between swabs and honey samples. Detection of RNA viral diversity was highest in honey samples, however, swabs also showed good recovery of plant and insect viruses. A minimum of two swab replicates was recommended to increase taxa yield for DNA metabarcoding. Our study suggests that swabbing of hive entrances is a highly effective biomonitoring method for non-invasively collecting hive eNA and delivering comprehensive surveillance for animal, plant and microbial communities in the environment.

Key words: Biosecurity, environmental biomonitoring, honey, non-invasive sampling, pathogens, pollen, surveillance, weeds

OPEN ACCESS

Academic editor:
Armando Espinosa Prieto
Received: 22 July 2025
Accepted: 13 September 2025
Published: 10 October 2025

Citation: Hill E, Milla L, Encinas-Viso F, O'Dwyer J, Gooden B, Hopper M, Roberts J (2025) Surface swabs outperform most traditional honeybee (*Apis mellifera*) hive samples for recovery of eDNA and eRNA. Metabarcoding and Metagenomics 9: e165436. https://doi.org/10.3897/mbmg.9.165436

Copyright: © Erin Hill et al.
This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Introduction

Sampling environmental nucleic acids (eNA), including DNA and RNA, has become a cornerstone of non-invasive monitoring of wildlife, from the detection of rare and elusive species and the characterisation of species assemblages within a given environment, to the surveillance of emerging biosecurity threats

(Valentini et al. 2016; Trujillo-González et al. 2019; Piggott et al. 2021). Much of this focus to date has revolved around the use of water samples for species monitoring, with standardised sampling techniques and protocols for a range of aquatic environments and detection applications (Pawlowski et al. 2020; Minamoto et al. 2021). While these methods are reliable for species wholly or partially dependent on aquatic environments for survival, alternative techniques have been proposed for the monitoring of terrestrial taxa, including soil, air, scats, and invertebrate pollinators (Banerjee et al. 2022; Van Der Heyde et al. 2022). Many of these techniques are still in their infancy and require thorough examination of potential advantages and limitations before being broadly applied for terrestrial biodiversity monitoring.

Genetic sampling from European honeybee (*Apis mellifera*) hives represents a promising repository for the detection of terrestrial taxa from a range of taxonomic sources. eNA methods have been used extensively in honeybee hives in the past, however, these applications have mostly assessed hive health through the detection of honeybee pests and pathogens (Ribière et al. 2019; Ribani et al. 2020, 2022; Biová et al. 2021), or identifying floral composition of honey to understand honeybee foraging requirements (Hawkins et al. 2015; Danner et al. 2017), and regional provenance of honeys being sold (Prosser and Hebert 2017; Milla et al. 2021; Wirta et al. 2021). More recently, attention has shifted to utilising honeybee hive samples for the identification of taxa unrelated to honeybee health and production outputs; these have included assessments of plant biodiversity and weed detection (Milla et al. 2022; Batchelor et al. 2023), identification of plant pathogens (Roberts et al. 2018b; Tremblay et al. 2019; Roberts et al. 2023), and detection of honeydew producing insects (Utzeri et al. 2018). These studies demonstrate the potential of honeybee hives to be jointly used for a range of eNA biosecurity and ecological monitoring applications that can both benefit the health and outputs of the given hive, as well as gaining an understanding of the surrounding environment and the incursion of new pests, weeds, and pathogens.

One of the challenges of using honeybee hives for biosecurity and biodiversity monitoring is the collection of appropriate samples and how this can be efficiently upscaled to allow thorough sampling of broad areas of interest. Commonly utilised samples such as honey, bees, and pollen require the assistance of specialists in the field to open and handle hives for these samples to be collected, which can be a limiting factor when undertaking large-scale studies. The collection of internal hive samples is also seasonally dependent as honey and pollen would not be found stored in sufficient amounts throughout the year and opening hives during colder months to collect samples can risk the survival of honeybee colonies. Additionally, sample collection of honeybee food sources (honey and pollen) can also weaken hives over time if long-term monitoring is undertaken. A need therefore exists to explore alternative options to sample eNA from honeybee hives that are easy for non-specialists to collect, are considered non-invasive for the ongoing persistence of hives and is applicable throughout the year.

A promising technique is the use of swabs to collect non-invasive samples of eNA from inside honeybee hives. To date, this method has been explored in the context of identifying taxa related to honeybee health, with varying success (Čukanová et al. 2023; Hénaff et al. 2023; Boardman et al. 2024; Mackay et al. 2025; Roberts et al. 2025). However, to fully understand the suitability of swabs for the collection and detection of taxa of biodiversity and biosecurity importance,

a comparison between the effectiveness of swabs to other commonly used hive sample types such as pollen, bees, and honey to detect a range of taxonomic groups is necessary. Furthermore, there is still little known about how these conventional sample types perform relative to each other and the potential implications of these differences on biomonitoring and surveillance outcomes.

Most eNA applications focus on capturing DNA for species detection, however there is a growing interest in how eRNA can be effectively collected and used in biomonitoring (Veilleux et al. 2021). RNA is expected to be less stable in the environment than DNA and may be able to provide information on local taxa that have been recently active in the environment (Yates et al. 2021; Scriver et al. 2023). Importantly, many viruses of interest for biosecurity surveillance are RNA viruses and therefore undetectable by eDNA methods. Honeybee RNA viruses can be important pathogens, and many studies have used samples of bees and hive products to investigate the hive virome (Roberts et al. 2018a; Kadlečková et al. 2022; Lee et al. 2023). Multiple plant RNA viruses are also significant crop pathogens, with many being pollen transmitted and commonly detected in hive samples (Fetters and Ashman 2023). However, there has been little examination of non-invasive methods to collect hive eRNA for virus detection (Čukanová et al. 2023).

The aims of this study were to: (1) test the performance of swab samples as an ideal non-invasive hive eNA collection method, and (2) compare swab samples to other hive sampling methods including honey, pollen, and bees. We did this by examining the taxonomic composition of DNA and RNA collected from swabs and hive derived samples including honey, pollen, and bees, coupled with a metabarcoding approach to identify bacterial, animal, fungal, and plant diversity, and a metagenomic approach to identify RNA viruses present within our chosen hive samples, to explore the full breadth of taxonomic diversity possible within a hive. Replicate samples were also collected for DNA metabarcoding to examine the sampling effort required to maximise taxonomic diversity detections. Sample types with the most diversity captured were identified for each taxonomic group, and optimal sampling strategies discussed for future biomonitoring and biosecurity surveillance efforts.

Methods

Study sites and sampling

Seven sample types were collected through the course of this study: hive entrance swabs, whole bees collected from inside the hive (top box bees), honey, internal pollen stored in pollen cells, hive bottom swabs (hive debris), pollen collected in hive entrance pollen traps, and returning forager bees. All seven of these sample types were compared for DNA metabarcoding (bacteria (16S), animal (COI), fungi (ITS1), plant (ITS2)), while only a subset of these sample types were used for the RNA virus comparisons (hive entrance swabs, internal pollen stored in pollen cells, top box bees, and honey). A summary of the samples collected for this study can be found in Table 1.

Each hive chosen for sampling comprised two boxes, were non-adjacent in the apiary where possible, and had stores of pollen and honey. All sampling and field tools were sterilised prior to each sampling event first

Table 1. Number of samples collected across each sample type and sampling day during this study. Values in parentheses represent the number of extraction replicates taken from each sample. Extraction replicates were not performed for any swab samples. RNA samples were pooled for sequencing.

	DNA samples				RNA samples	
Sample type	**Day 1**	**Day 2**	**Day 3**	**Total samples**	**Day 1**	**Total samples**
Hive entrance swab	9	9	9	27	6	3
Hive debris swab	9	9	9	27	-	-
Honey	-	-	9 (×2)	18	3 (×2)	3
Internal pollen	-	-	6 (×2)	12	3 (×2)	3
Pollen trap	3 (×2)	3 (×2)	3 (×2)	18	-	-
Return forager bees	3 (×2)	3 (×2)	3 (×2)	18	-	-
Top box bees	3 (×2)	3 (×2)	3 (×2)	18	3 (×2)	3

using 10% bleach solution, followed by deionised H_2O, and then 70% ethanol. Equipment used between hives (e.g. hive tool, tweezers, micro spatulas) was cleaned with 70% ethanol in the field prior to attending each hive. All samples collected during each trip were immediately placed on ice in the field, then stored at -80 °C until DNA/RNA extraction. No storage buffers were used for any of the samples collected.

DNA sampling

Hive samples collected for metabarcoding analysis were located at the Gungahlin Homestead in Crace, Australian Capital Territory (ACT) (-35.223300°S, 149.121971°E), which at the time of this study was a former pastoral property occupied by the Commonwealth Scientific and Industrial Research Organisation (CSIRO). The site is surrounded by grassland nature reserves to the north and south, and residential areas to the east and west of the property. The apiary at this site comprised 12 hives total, of which three hives were utilised for this study. Sampling was undertaken on the 20th, 24th, and 27th of January 2023, with sample collections split into morning and afternoon sessions. Hives were sampled in a different order each trip to minimise any timing effects, particularly from pollen trapping.

In the morning sampling session, the hive entrance was swabbed using a sterile nylon FLOQSwab (Copan Diagnostics Inc.) dipped in deionised H_2O. Each swab was rubbed across the entire length of the hive entrance four times, twisting it throughout, and attempting to reach a range of depths inside the hive. Swabbing of the hive entrance was repeated two more times with a new swab each time. The hive was then opened, and bees present in the top box on an outer honey frame were collected in a sterile 50 mL falcon tube, collecting approximately 50 bees. Following this, the top box and frames from the bottom box were removed to allow swabbing of the bottom of the hive. Swabbing the bottom of the hive was carried out as described above, avoiding any remaining bees and concentrating on areas where hive entrance swabs were unlikely to have reached. Three consecutive swabs were also collected for the bottom of the hive. The hive was then re-assembled, and a sterilised pollen trap was attached to the front of the hive to collect pollen from the hind-legs of returning forager bees. The time the pollen trap was attached was recorded.

On the final sampling trip during the morning session, samples of honey and internal pollen were also collected from each hive prior to swabbing the hive bottom. These samples were only collected on the final sampling day due to their presumed long-storage period, therefore unlikely to show significant day-to-day variation. Three honey samples were collected, all from the top box but across separate frames. Approximately 50 mL of honey was collected in a 100 mL specimen collection container from each frame. Internal pollen was similarly collected across two frames; a total of 10 pollen cells were sampled from each frame, by scooping out the contents of individual pollen cells using a sterile micro spatula and placing into a 100 mL specimen collection container for each frame.

In the afternoon sampling session, the pollen collection tray underneath the pollen trap was removed and the time noted, and pollen collected into a 50 mL falcon tube. Throughout the experiment, pollen traps were closed for approximately 5 hours during the day to allow for pollen collection. Pollen traps were left open overnight to reduce impact on the hives. Forager bees returning to the hive with visible pollen on their hindlegs were then collected with sterile tweezers and placed into a 50 mL falcon tube, with the aim to collect 20 returning foragers per hive, however, this was not always possible. On the second sampling trip for this study, a pollen trap was improperly fitted to one of the hives, resulting in no pollen present in the trap in the afternoon session. To ensure a sample was still collected from this hive at roughly the same time as the other hives, the pollen trap was repositioned and left on the hive overnight, with the pollen collected at midday the following day. Returning forager bees were still collected from this hive at the intended afternoon sampling session.

RNA sampling

Hive samples collected for RNA sequencing were located at the CSIRO Black Mountain site in Acton, ACT (-35.275051°S, 149.112159°E). This site borders Black Mountain and the Australian National Botanic Gardens to the west, the Australian National University immediately east, and residential areas to the north and south of the property. The apiary at this site comprised a total of five hives, of which three hives were sampled. RNA sampling was undertaken in a single session on the 11th of April 2024, with sampling methods and order the same as above for the DNA samples, however, only two swab replicates were collected for RNA analysis. Samples taken for RNA analysis included hive entrance swabs, top box bees, internal pollen cells and honey.

Sample processing

Honey, pollen, and bee samples required pre-processing to produce viable samples for DNA and RNA extractions, including a bead beating step to ensure samples were homogenised and allow lysis of tough pollen, bacterial, and fungal particles that may have been present in the samples.

Honey samples collected from each hive were aliquoted into approximately 10 mL volumes and each sample was weighed. Two replicates from each frame were then chosen for extractions, with equal starting material weights considered when choosing replicates. Ultrapure H_2O was then added to each sample to create final volumes of 50 mL and incubated at 40 °C for 30 minutes with regular mixing

to homogenise the solution and dissolve sugars present. Samples were then centrifuged at 5000 rpm for 20 minutes and the supernatant discarded. Remaining pellets were then resuspended in 500 µL of 1 × PBS buffer and transferred to a 2 mL BeadBug™ homogenizer tube containing 1.5 mm zirconium beads.

All pollen samples were weighed and homogenised prior to subsampling for extractions. For the internal pollen samples, 5 mL of 1 × PBS buffer was added. Pollen trap samples were weighed and homogenised using a mortar and pestle, and 1 g of pollen subsampled and 5 mL of 1 × PBS buffer added. Some pollen trap samples weighed < 1 g so the entire sample was used, however, volumes of 1 × PBS buffer were adjusted to maintain a similar pollen:PBS buffer ratio to the other pollen trap samples. Samples were then homogenised further using a thermomixer set at room temperature for 1 hour at 1000 rpm for pollen frames and pollen trap samples. Following this, 500 µL of solution was aliquoted into a 2 mL BeadBug™ homogenizer tube containing 1.5 mm zirconium beads. Two replicates were taken from each of the pollen trap and internal pollen frame samples.

The number of return forager bees per sample was counted and placed into an extraction bag (Seward) for homogenisation. Care was taken to ensure pollen did not detach from the bees' legs during this process. Approximately 10 mL of PBS buffer was then added to each bag, and the mixture was initially crushed using a pestle on the outside of the strainer bag. Following this, strainer bags were passed through a pasta rolling machine (Bialetti) multiple times to ensure bees were thoroughly homogenised. The supernatant was then collected, and 500 µL of solution was transferred to a 2 mL BeadBug™ homogenizer tube containing 1.5 mm zirconium beads. This process was repeated for the top box bees, ensuring an equivalent number of top box bees were used in comparison to the return forager bees collected from the same hive/sampling trip. For the RNA samples, 20 top box bees were used as starting material. Two replicates were taken from each of the homogenised bee samples.

BeadBug tubes containing sample and 500 µL of 1 × PBS buffer were centrifuged at 13,000 rpm for 1 minute to pellet each sample, and the supernatant discarded. 550 µL of buffer CF (Macherey-Nagel Nucleospin Food Kit) was then added for the DNA samples, and 500 µL of Lysis Buffer (Maxwell RSC simply RNA Tissue Kit (Promega)) was added for the RNA samples. Bead beating was performed using a TissueLyser (Qiagen), comprising two rounds of bead beating at 30 Hz for 90 seconds.

DNA extractions and amplification

DNA extractions of honey, pollen, and bee samples were performed using the Macherey-Nagel Nucleospin Food Kit following the protocols described above, following the manufacturer's instructions. Samples were left to sit at room temperature for 1 hour in a final elution of 100 µL of Buffer CE before the last centrifugation step. Swab samples were extracted using a QIAamp DNA Investigator Kit (Qiagen) following the protocol for isolation of total DNA from surface and buccal swabs. The procedure was followed for the use of cotton or Dacron swabs, with the optional carrier RNA and QIA shredder steps included. Samples were left to sit at room temperature for 1 hour in a final elution of 100 µL of Buffer ATE before the last centrifugation step.

Table 2. Metabarcoding primers used for amplification in this study. PNA clamp sequence developed for this study to reduce *Apis mellifera* amplification is listed with the animal (COI) primers.

Region	Primer Name	Sequence (5'-3')	Target Taxa	Reference
16S	515-FY	GTGYCAGCMGCCGCGGTAA	Bacteria	(Apprill et al. 2015; Parada et al. 2016)
	806R	GGACTACNVGGGTWTCTAAT		
COI	FwhF2	GGDACWGGWTGAACWGTWTAYCCHCC	Animal	(Vamos et al. 2017)
	FwhR2n	GTRATWGCHCCDGCTARWACWGG		
	Amel_PNA	CATTCTTCACCTTCAGTAGA		
ITS1	ITS1f	CTTGGTCATTTAGAGGAAGTAA	Fungi	(White et al. 1990; Gardes and Bruns 1993)
	ITS2r	GCTGCGTTCTTCATCGATGC		
ITS2	S2F	ATGCGATACTTGGTGTGAAT	Plant	(Chen et al. 2010)
	4rev	TCCTCCGCTTATTGATATGC		

Negative extraction controls were run throughout the DNA extraction process for all sample types. All samples were quantified using a Nanodrop 8000 Spectrophotometer (Thermo Scientific) and then normalised to 5 ng/µL for PCR amplification to ensure differences in OTU composition were not affected by starting concentrations of samples. Any samples with an initial concentration < 5 ng/µL were kept at a neat concentration.

Samples were amplified using four universal primer pairs targeting bacterial (16S), animal (COI), fungal (ITS1), and plant (ITS2) species (Table 2). PCRs were carried out in 15 µL reactions, containing 7.5 µL of GoTaq Green Master Mix (Promega), 0.6 µL each of forward and reverse primers (10 uM), 4.3 µL of ultrapure H_2O, and 2 µL of DNA. PCR amplification comprised an initial denaturation at 95 °C for 5 minutes, followed by 30 cycles of 95 °C for 30 seconds, 50 °C for 30 seconds, 72 °C at 50 seconds, and a final extension of 72 °C for 5 minutes. The cycle number was increased to 35 cycles for the amplification of the fungal ITS1 and animal COI primers. Animal COI reactions included a PNA clamp designed to reduce *Apis mellifera* amplification (Suppl. material 1). Modifications to the protocol above included the addition of 0.75 µL of PNA clamp (20 uM) to the reaction and reduction of ultrapure H_2O to 3.55 µL, the use of GoTaq Enviro qPCR System (Promega) in lieu of the Taq described above to improve PCR amplification, and the initial denaturation step changed to 2 minutes as recommended for the GoTaq Enviro qPCR System. Negative PCR controls were run throughout the amplification process, and a subsample of PCR products were visualised using a 2% agarose gel to confirm successful amplification. Three replicate PCRs were completed for each marker and then pooled at equal amounts for a final volume of 30 µL. PCR products were then quantified using a Fluoroskan (Thermo Fisher Scientific) and a Quant-iT dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific), and all samples/markers normalised to 10 ng/µL. Samples were then pooled at equal amounts across the four markers, and sent to the Biomolecular Resource Facility in Canberra, Australia, for indexing and Illumina MiSeq sequencing using a v3 600 cycle kit.

RNA extractions

RNA extractions of all sample types were performed using the Maxwell RSC simplyRNA Tissue Kit (Promega). For swab samples, swab tips were added to a 1.5 mL Eppendorf tube containing 500 µL of Lysis Buffer (Promega) and placed

on a thermomixer set for shaking at 900 rpm for 1 hour at room temperature. Following bead beating of the honey, pollen, and bee samples and shaking of swab samples, 350 µL of lysate was collected and added into a 1.5 mL Eppendorf tube containing 350 µL of Lysis Buffer. Samples were then prepared for use on a Maxwell RSC Instrument, for automated RNA extraction. Two extraction replicates were undertaken for each hive sample, and negative extraction controls were run throughout the RNA extraction process. Extraction replicates were then pooled by sample type and hive and quantified using a 4200 TapeStation System (Agilent). Samples were then sent to the Biomolecular Resource Facility (BRF) in Canberra, Australia, for library preparation with the Illumina Stranded mRNA Prep kit and sequencing using a NextSeq 2000 P2 2 × 150 bp run.

Metabarcoding analysis

The demultiplexed paired FASTQ reads were quality trimmed to Q20 using bbduk (Bushnell 2014) and merged using USEARCH. Gene-specific primer sequences were used to split reads into markers using cutadapt (Martin 2011) and low-quality reads were filtered out using the USEARCH fastqfilter function. For each marker, sequences were denoised and zero-radius operational taxonomic units (ZOTUs) were chosen using the USEARCH unoise3 function using a minimum cluster size of 8 reads. For bacterial 16S, we continued to assign taxonomy to ZOTUs. For the other three markers, we clustered in 97% identity OTUs using USEARCH cluster_fast function.

To assign taxonomy to predicted ZOTU/OTUs, we used the USEARCH sintax function with an 90% bootstrap cutoff against the corresponding marker databases. The sintax command uses the kmer-based SINTAX algorithm (Edgar 2016) to make taxonomic predictions. For animal COI, we downloaded the GB255 version of the SINTAX-formatted MIDORI2 database (Leray et al. 2022) from the MIDORI website (www.reference-midori.info). For plant ITS2, we downloaded full *Streptophytina* sequences from the ITS2DB (Selig et al. 2007; Koetschan et al. 2010, 2012; Merget et al. 2012; Ankenbrand et al. 2015) website (http://its2.bioapps.biozentrum.uni-wuerzburg.de) on December 2023. We used taxonkit (https://bioinf.shenwei.me/taxonkit) to add taxonomic lineage information and reformat the FASTA header to SINTAX format. For fungal ITS1, we downloaded the SINTAX-formatted file utax_reference_dataset_25.07.2023.fasta.gz from the UNITE v9.0 (Abarenkov et al. 2024) database website (https://unite.ut.ee/repository.php). For bacterial 16S, we downloaded the RDP training set v19 from the RDP website (https://sourceforge.net/projects/rdp-classifier/files/RDP_Classifier_TrainingData) and reformatted it to SINTAX format using a custom python script. We generated searchable databases for all markers with the USEARCH v11.0.667_i86linux32 (Edgar 2010) makeudb program. Secondary identifications were assigned using blastn against the full NCBI nt database, returning the top ten matches. The top BLAST results were parsed using the python script taxonomy_assignment_BLAST_V2.py (github: Joseph7e/Assign-Taxonomy-with-BLAST), which assigns the consensus lowest taxonomic level possible from all hits that match the provided identity percent cut-offs and e-value. We assigned the consensus BLAST results with a minimum of 97% identity and e-value of $1e^{-10}$ to missing taxonomic assignments not found in marker-specific databases.

Once ZOTU/OTUs were identified, we used the R package microDecon (McKnight et al. 2019) to reduce contaminant reads or remove ZOTU/OTUs that were abundant in negative controls. Additionally, we excluded COI identifications that were assigned to phyla Ascomycota and Basidiomycota, 16S identifications that were not bacteria or were assigned to chloroplast or mitochondria, and ITS2 identifications that were not assigned to phyla Streptophyta (ITS2DB) or Viridiplantae (NCBI). Finally, we used the rarefy function of the R package vegan (Oksanen et al. 2019) to generate rarefaction curves and identify samples with 100 or fewer reads and excluded them from further analysis.

From the filtered dataset, we carried out the following comparisons: (a) taxonomic coverage across sample types; and (b) effect of replicates (sample and DNA extraction) on taxonomic coverage. To identify taxonomic coverage across sample types, the number of observed taxa was plotted for each sample type and metabarcoding marker. To keep sample replication consistent across all sample types, only the first two swab replicates were included in this analysis. Any significant differences between number of taxa detected across sample types were noted for each marker using t-tests, performed using the R package rstatix 0.7.2 (Kassambara 2019). The top 10 genera present in the most samples across all sample types were then identified for each taxonomic group and plotted to identify consistency of common genera across sample types. NMDS was calculated using vegan for each marker to identify if taxonomic compositions differed between sample types. Analysis of multivariate homogeneity of group dispersions (variances) was calculated from NMDS results using the betadisper function in vegan, and significant dispersion differences identified using ANOVA. Shannon diversity indices were also calculated to estimate the richness and evenness of sampling for each marker and sample type using the estimate_richness function of the R phyloseq package (McMurdie and Holmes 2013). Species accumulation curves were calculated for each marker and sample type using the specaccum function in vegan to identify the sampling effort necessary for appropriate taxonomic coverage for each sample type. We ran a generalised linear model with gamma distribution to explore the effects of variables including individual hives, sample type, collection date, and combinations of these, on the diversity of taxa detected. Analysis of deviance was then calculated using the Anova function of the R car package (Fox and Weisberg 2018) across all sample types for bacteria (16S), fungi (ITS1), and plants (ITS2), to identify variables that may have significantly influenced the diversity of taxa detected using the Shannon diversity index.

To assess the effects of replicates on taxonomic recovery, the percentage overlap of taxa detected across extraction replicates was calculated to identify how similar each replicate was in taxonomic composition. For this analysis, only the first two swab replicates were retained to provide consistency when comparing percentage overlap with the other sample types. Following this, more in-depth analysis of all three swab replicates was performed. The number of unique taxa observed in each swab replicate was identified to confirm if subsequent swabbing resulted in fewer detections of unique taxa. This was expected as swab replicates sampled the same surfaces in the hives and were predicted to remove DNA during the swabbing process, leaving less deposited DNA available for sampling in subsequent swabbing replicates.

RNA sequencing analysis

Returned raw reads were filtered, assembled and classified using the MetaDIVE pipeline (https://github.com/James-ODwyer/MetaDIVE/). Briefly, reads were filtered and adaptors trimmed using fastp v.0.23.2 (Chen et al. 2018) with the following settings: minimum length 85 bp, individual base minimum quality filter of 16, complexity threshold of 10, a 4 bp front sliding window, minimum quality window of 20, and average minimum quality filter of 19. Reads were then assembled using trinity v.2.8.5 (Grabherr et al. 2011) with a minimum assembly size of 400 bp and contigs were taxonomically classified using a two-step process of Diamond BLASTx v.2.0.15 (Buchfink et al. 2021) and Kraken v.2.1.3 (Wood et al. 2019) using the preset setting "--mid sensitivity" for diamond, and BLASTn v.2.13.0 (Camacho et al. 2009) using the settings "-word_size 25" against NCBI's non-redundant nt and nr databases (updated September 2024).

From the outputs generated by MetaDIVE additional filtering of identified viral contigs was performed in R. Contigs which matched the following criteria were removed from the dataset: a mean sequence complexity lower than 35, a maximum percent identity < 85, and a contig length < 500 bp. Any uncultured viruses or viruses unable to be assigned to species were also filtered from the dataset. The dataset was also checked for the presence of any phages as these groups are often poorly assigned; no phages were identified in the dataset. Finally, viruses that appeared in the negative control were down sampled in the experimental samples based on number of reads assigned to each contig.

From the filtered dataset, only comparisons of taxonomic coverage across sample types could be investigated due to the absence of sequenced RNA replicates. To identify taxonomic coverage across sample types, the number of observed viruses were plotted for each sample type. The top 10 viruses present in the most samples across all sample types was plotted to identify consistency of common RNA viruses across sample types. Shannon diversity indices were also calculated and plotted for each sample type using the plot_richness function of the R package phyloseq. Viruses were then classified according to their host species (animal, fungi, honeybee, insect, or plant) and plotted using ggplot2 3.5.2 (Wickham 2016) to identify if sample types were associated with the recovery of virus types.

Results

Metabarcoding

After quality control and filtering steps, we obtained a total of 5,273,034 metabarcoding reads split across the four markers (Suppl. material 2: table S1). We merged all ZOTUs to their lowest taxonomic identification, and from herein we use the terms taxa/taxon to refer to these assignments.

A total of 292,078 reads was returned for bacteria (16S), 138,503 reads returned for animal (COI), 1,402,782 reads returned for fungi (ITS1), and 3,439,671 reads were returned for plant (ITS2) samples. In the bacteria (16S) dataset, the top box bees produced the highest number of total reads (133,221) and average number of reads per sample (7401.17 ± 3788.68 S.D.), however the hive debris swabs and hive entrance swabs recovered the highest number of unique taxa (33 each out of 35 total observed), and the hive debris swab samples

additionally comprised the highest number of average taxa detected per sample (22.04 ± S.D.). The pollen trap samples produced the lowest number of total reads (2016) and total unique taxa (23), while the internal pollen samples comprised the lowest number of mean reads per sample (70.69 ± 111.05 S.D.) and mean unique taxa per sample (5.4 ± 3.11 S.D.). In the animal (COI) dataset, the hive entrance swabs returned the highest number of total read (52,906), mean reads per sample (2300.26 ± 3999.39 S.D.), and mean unique taxa per sample (2.83 ± 0.83 S.D.), while the internal pollen samples was lowest for these calculations (937 total reads, 58.56 ± 120 S.D. average reads per sample, and 1.5 ± 0.73 S.D. mean unique taxa per sample). However, all substrates detected all four unique taxa observed in this dataset. For the fungi (ITS1) dataset, the hive entrance swab samples again returned the highest number of total reads (347,253), average reads per sample (13355.88 ± 6983.56 S.D.), and mean unique reads per sample (56.58 ± 12.33 S.D.), and additionally detected the highest number of total unique taxa, 232 of the 309 total fungal taxa observed in this dataset. The honey samples comprised the lowest number of total reads (42,990) and mean reads per sample (2528.82 ± 1537.82 S.D.), however, the top box bee samples returned the lowest number of total unique taxa (90) and mean taxa observed per sample (13.33 ± 4.81 S.D.) of any of the substrates for this dataset. In the plant (ITS2) dataset, the internal pollen samples returned the highest value for total reads (1,038,153), however, was lowest for average number of unique taxa observed per sample (12.9 ± 3.16 S.D.). The honey samples produced the highest average number of unique taxa observed per sample (23.76 ± 8.4 S.D.), however, returned the lowest values of total reads (187,170) and mean reads per sample (11,010 ± 9130.55 S.D.) of any of the substrates for this dataset. The return forager bees produced the highest mean number of reads per sample (29,206.39 ± 13,987.88 S.D.), however, identified the lowest number of taxa, with 33 of the 62 total plant taxa detected. The hive debris swabs detected the most unique plant taxa, with 56 taxa total observed from this substrate.

Taxonomic coverage across sample types

Across the bacterial (16S), animal (COI), and fungal (ITS1) metabarcoding markers, swab samples from the hive entrance and hive debris were consistently shown to comprise the highest number of taxa detected of all sample types (Fig. 1). Plant (ITS2) taxa detections were highest in the honey samples, closely followed by the swab samples, however no significant differences in taxa detections were observed between these three sample types (Suppl. material 2: table S2). Pollen trap and internal pollen samples yielded fewer detections for bacteria (16S) and animals (COI) relative to the other sample types collected, with most comparisons appearing significantly different between sample types for the bacteria (16S) (Suppl. material 2: table S3). Internal pollen samples were particularly poor at detecting animal (COI) taxa, with individual samples commonly only detecting a single taxon. A high degree of variability was observed between sample types amplified for the fungi (ITS1), with most comparisons between sample types being significantly different (Suppl. material 2: table S2).

Analysis of the top ten genera present in all samples revealed most of these taxa were present in all sample types for each marker (Fig. 2). The bacterial (16S) dataset showed the most discrepancies between sample types; both hive swab

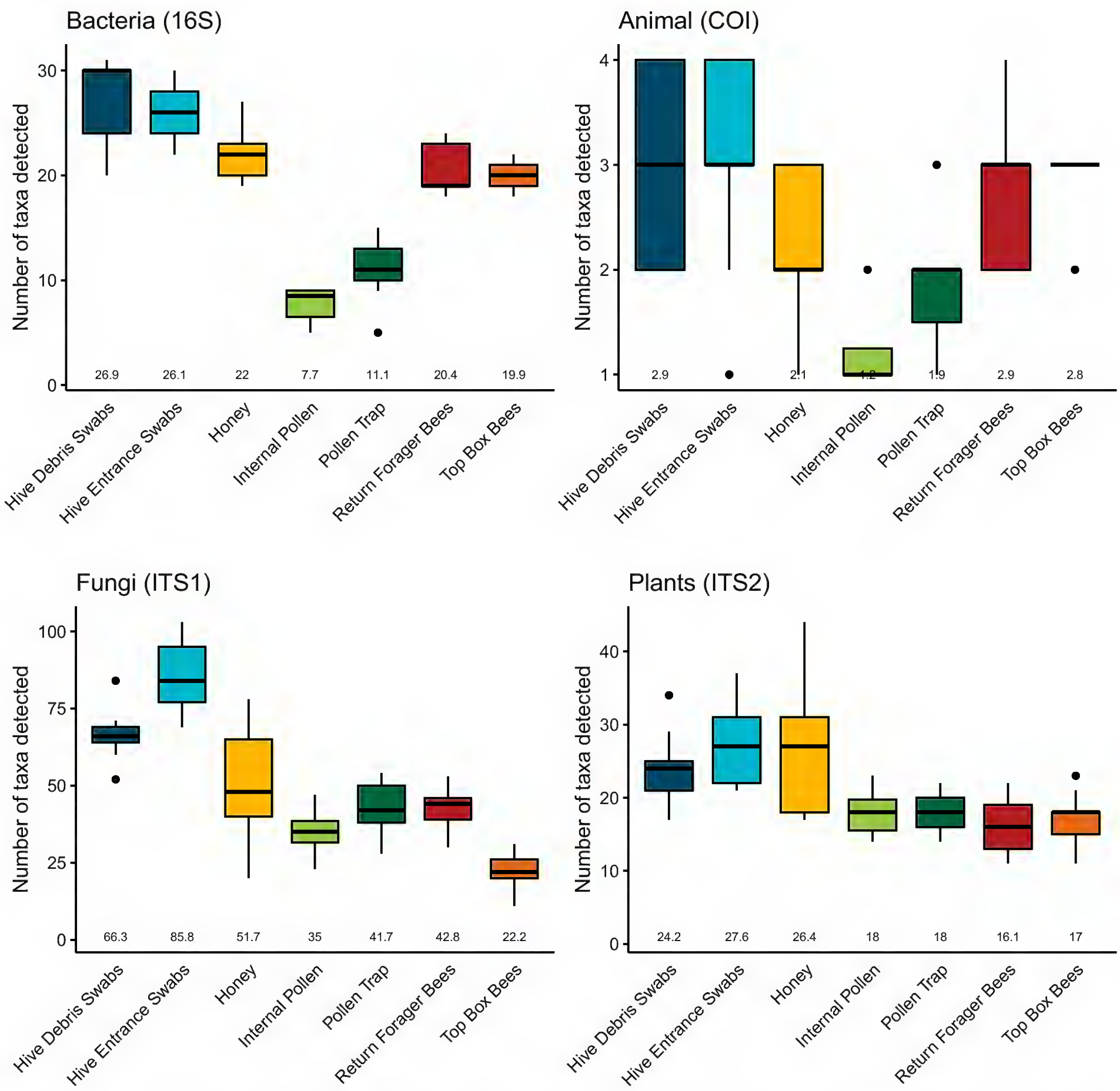

Figure 1. Number of taxa detected across each sample type for bacteria (16S), animal (COI), fungi (ITS1) and plant (ITS2) metabarcoding markers. Mean values for each substrate and marker are shown below each boxplot.

sample types and honey detect all ten taxa, while both bee sample types failed to detect *Bradyrhizobium* sp. In the pollen samples, only seven of the ten bacterial (16S) taxa were present in the internal pollen, and eight of ten were present in the pollen trap samples. In the fungal (ITS1) dataset, the genus *Epicoccum* was not observed in either of the swab sample types; otherwise, all ten genera were identified in all sample types. In the plant (ITS2) dataset, all ten genera were present in each sample type except for *Kunzea* sp. in the pollen trap samples. All sample types detected all genera present in the animal (COI) dataset.

NMDS analysis showed a high degree of overlap across sample types and marker, however, with some differentiation in the bacteria (16S) and fungi (ITS1) datasets (Fig. 3). The internal pollen and pollen trap samples appear to be compositionally different to the other sample types in the bacteria (16S) dataset, while there appears to be a gradient of compositional differences in the fungi

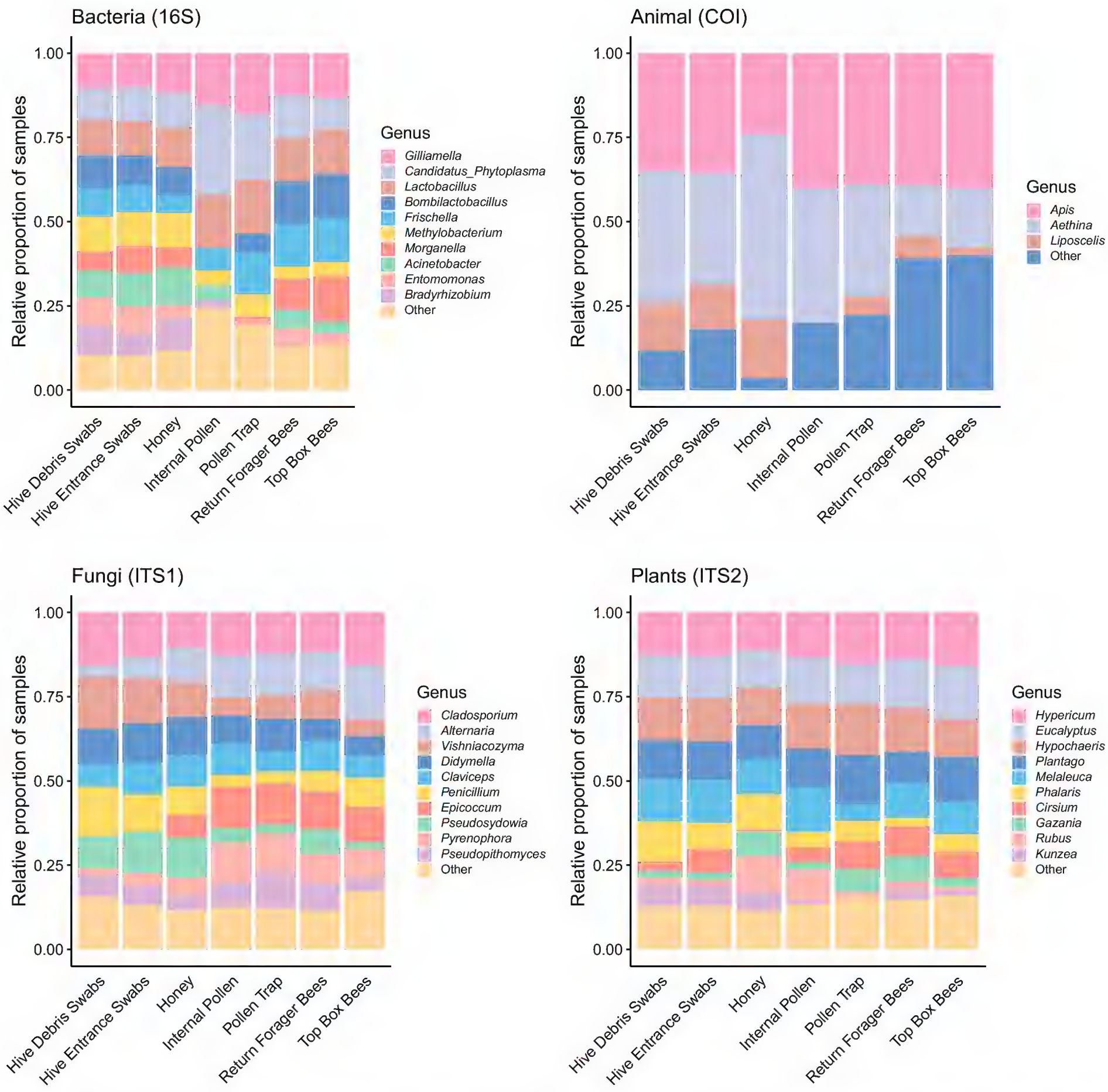

Figure 2. Top 10 genera detected across all samples, and their detection proportions across each sample type for bacteria (16S), animal (COI), fungi (ITS1) and plant (ITS2) metabarcoding markers.

(ITS1) dataset, with hive debris swabs and hive entrance swabs towards one end, and internal pollen, pollen trap, and top box bees towards the other end. The plant (ITS2) dataset does not appear to show any distinct clustering by sample type, and while the animal (COI) dataset does show some clustering, this does not appear to be influenced by sample type and is more likely an effect of the low number of taxa detected using this marker. Analysis of multivariate homogeneity of group dispersions were significant for all markers (bacteria (16S) $F(6, 112) = 18.88$, $p < 0.001$; animal (COI) $F(6, 95) = 4.60$, $p < 0.001$; fungi (ITS1) $F(6, 112) = 9.87$, $p < 0.001$; plant (ITS2) $F(6, 112) = 4.47$, $p < 0.001$).

Shannon diversity indices indicate that honey, hive entrance, and hive debris swabs comprise similar alpha diversity measures in the bacteria (16S) and plant (ITS2) datasets (Suppl. material 2: fig. S1). In the animal (COI) dataset, hive entrance swabs appear to show the highest levels of alpha diversity, while

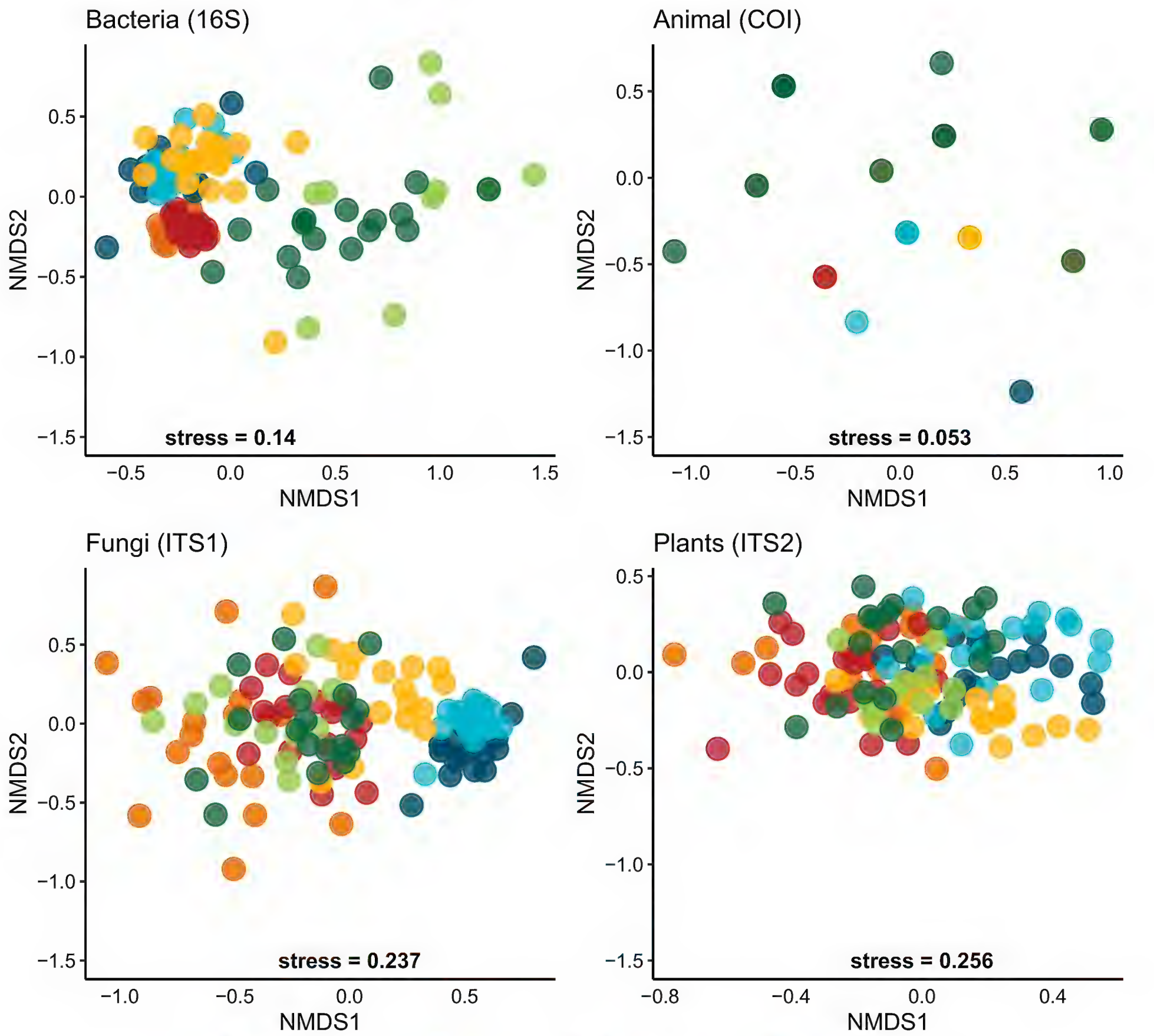

Figure 3. NMDS plots examining compositional differences between sample types for bacteria (16S), animal (COI), fungi (ITS1) and plant (ITS2) metabarcoding markers.

honey samples outperform the swab samples in the fungal (ITS1) dataset. Shannon diversity indices consider both the richness of a sample (i.e. number of taxa observed), and the evenness of the sampling (i.e. abundance of each species), indicating that honey samples may be detecting fungal (ITS1) taxa more consistently across samples than other sample types.

Species accumulation curves show that the swab samples consistently outperform the other sample types in detecting taxa, with hive entrance swabs returning the highest number of taxa to sampling effort for animal (COI), fungi (ITS1), and plant (ITS2) markers, and hive debris swabs highest for the bacteria (16S) (Suppl. material 2: fig. S2). Accumulation curves for the bacteria (16S) and animal (COI) datasets show that only four samples are necessary from both swab sample types to reach an asymptote. Accumulation curves however

Table 3. Analysis of deviance (type II tests) calculated for bacteria (16S), fungi (ITS1) and plant (ITS2) metabarcoding markers, using the Shannon diversity response variable. Df = degrees of freedom; X^2 = Chi-Square value. Significant values are denoted with an asterisk (*).

		Bacteria (16S)		Fungi (ITS1)		Plants (ITS2)	
Variable	Df	X^2	p-value	X^2	p-value	X^2	p-value
Hive	2	2.95	0.23	3.94	0.14	0.16	0.92
Sample type	6	29.14	5.73e-05*	168.1	2.2e-16*	24.21	0.0005*
Collection date	3	0.24	0.97	9.31	0.03*	2.14	0.54
Hive + sample type	12	9.79	0.63	20.12	0.06	16.68	0.16
Hive + collection date	4	1.05	0.9	8.97	0.06	4.99	0.29
Sample type + collection date	8	1	1	34.12	3.86e-05*	4.93	0.76
All variables	15	1.04	1	26.82	0.03*	16.9	0.32

did not reach an asymptote for the fungi (ITS1) or plants (ITS2) for any of the sample types, indicating a need to increase sampling effort for these groups.

Analysis of deviance across all sample types for bacteria (16S), fungi (ITS1), and plants (ITS2) shows that for each of these markers, sample type was a significant variable that influenced the diversity of taxa detected (Table 3). Additionally, sample type and collection date individually, sample type and collection date as combined variables, and all variables considered (hive, sample type, and collection date), were significant contributors to the variances detected in the fungal (ITS1) dataset.

Effect of replicates on taxonomic coverage

There was considerable variation in the taxa detected between extraction and sample collection replicates across the sample types (Fig. 4). Mean percentage overlap of taxa detected appeared lowest overall for the fungi (ITS1), ranging from 19.1% in the top box bee samples to 40.6%in the hive entrance swabs, indicating high variability in taxa detected across replicates. In the bacteria (16S), variation appeared to be influenced by sample type, with bee samples (top box bees and return forager bees) showing the highest average percent overlap across extraction replicates (86.5% and 83.8% respectively), followed by swab samples (hive entrance 69.8% and hive debris 68.8%) and honey (61.7%), and pollen samples showing the lowest percentage overlaps of 39.% for internal pollen and 43.3% for pollen trap. Overlap was moderately consistent across replicates for each sample type in the plant (ITS2) dataset (range 58% to 77.9%), however slightly higher in the pollen and honey samples. Average percent replicates overlap in the animal (COI) taxa detected was highest in the bee samples and lowest in the pollen samples, however, only four taxa total were identified in this dataset. An overlap of 0 was identified for the internal pollen samples sequenced using animal (COI); this has occurred as only one sample was successfully sequenced for multiple replicates, and each of these replicates amplified a different taxon.

When comparing the number of unique taxa collected from each swab replicate a downward trend was observed in the return of unique taxa following subsequent swabbing of the same area (Fig. 5). For the bacteria (16S) only one additional unique taxon was detected from a second swab replicate in the hive entrance swab samples, with no new taxa detected from either swabbing

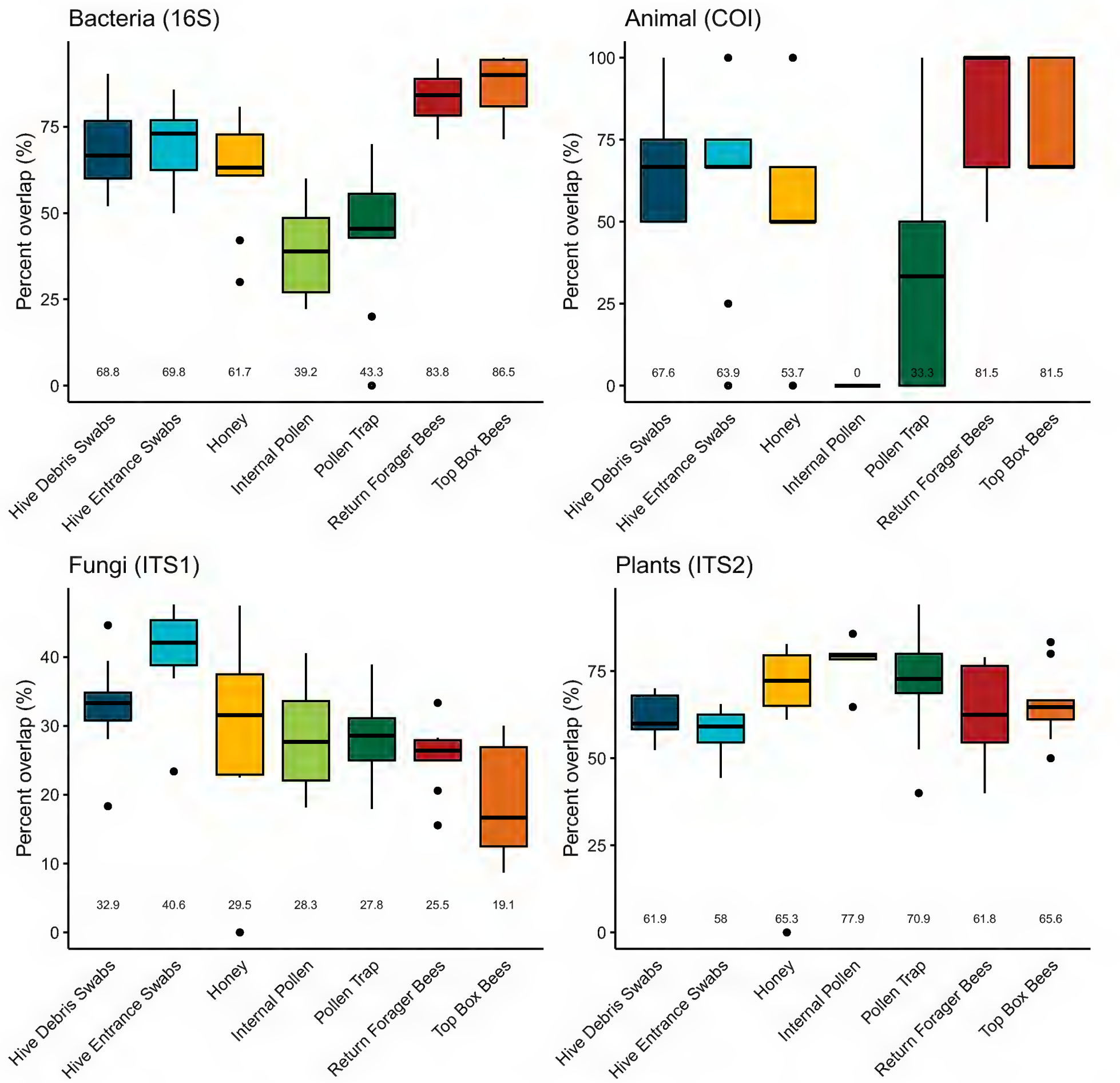

Figure 4. Percentage overlap of taxa detected across DNA collection and extraction replicates for each of the sample types sequenced for bacteria (16S), animal (COI), fungi (ITS1) and plant (ITS2) metabarcoding markers. Mean overlap percentages are shown below each boxplot. For hive entrance and debris swabs only the first two replicates were compared for consistency with the other sample types.

technique on the third replicate, and no new taxa appearing from a second hive debris swab replicate. In the animal (COI) dataset, all four taxa were detected from first swabs of the hive entrance and hive debris. The fungi (ITS1) swabbing replicates continued to detect additional unique taxa across each replicate for both the hive entrance and the hive debris, with 31 and 19 additional taxa detected in replicates two and three of the hive entrance swabs, and 42 and 17 additional taxa in replicates two and three of the hive debris samples. In the plant (ITS2) dataset, replicates one and two detected most of the observed taxa for both the hive entrance swabs and the hive debris swabs, with seven additional unique taxa detected in the second swab replicate for the hive entrance

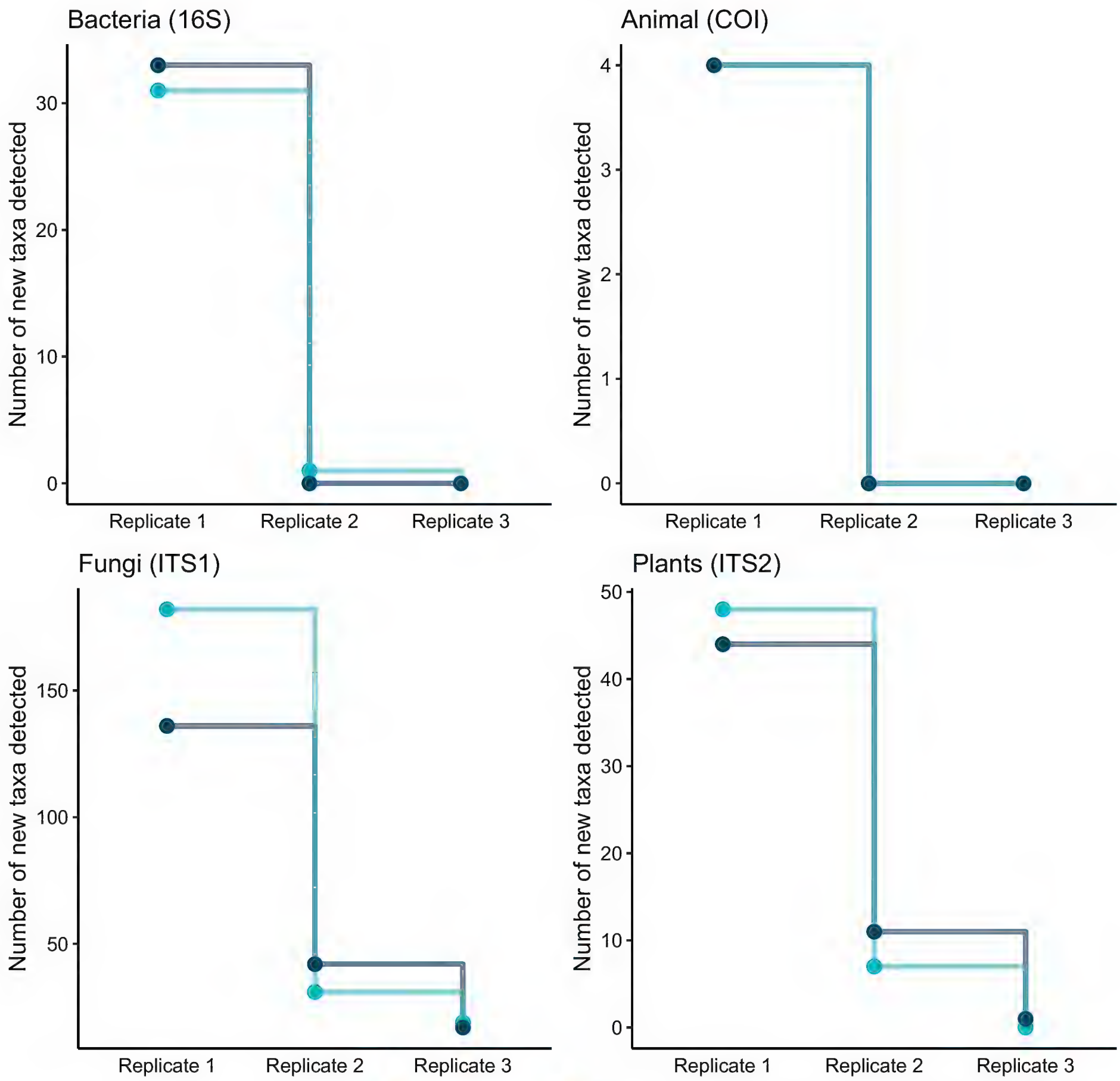

Figure 5. Step plots showing the number of taxa detected by the first swab replicates taken from the hive entrance and hive debris, and the number of additional unique taxa detected by performing additional swab replicates for bacteria (16S), animal (COI), fungi (ITS1) and plant (ITS2) metabarcoding markers.

samples and 11 unique taxa detected in the second swab replicate of the hive debris samples. Only one unique taxon was detected in the third swab replicate for the hive debris samples in this dataset, and no additional taxa in the third swab replicate for the hive entrance swab samples.

RNA sequencing

RNA sequencing resulted in 872 million raw reads returned across 13 samples. Raw assignments showed that most reads were assigned to eukaryote species in the hive entrance swabs, internal pollen, and top box bee samples, including *Apis mellifera*, particularly in the top box bee samples (Fig. 6).

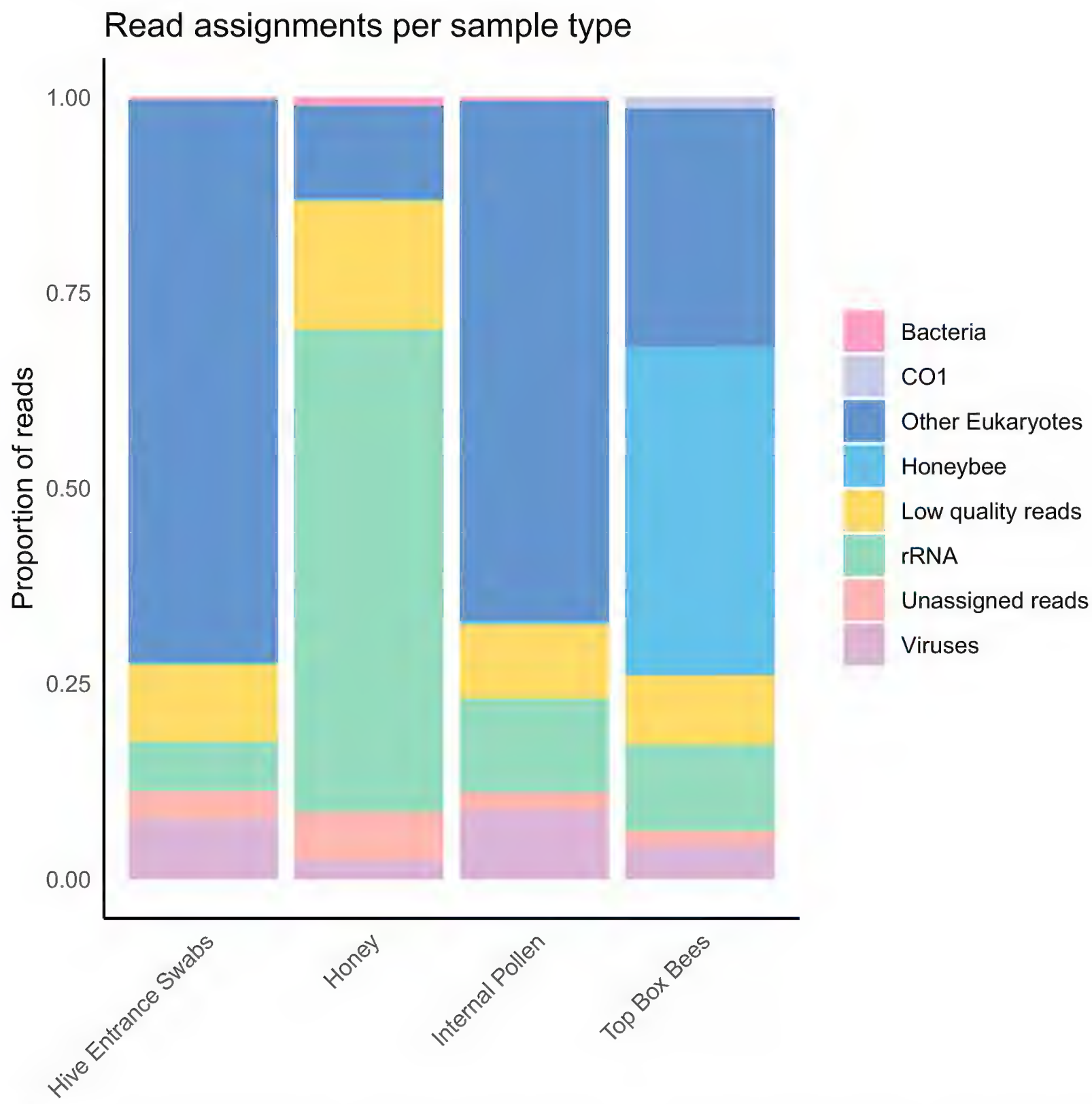

Figure 6. RNA raw read outputs and the proportions assigned to different taxonomic groups for each sample type.

The top 10 eukaryote species (excluding *A. mellifera*) assigned in this group were predominantly plant species, with a single fungal species (*Aureobasidium melanogenum*) detected in the hive entrance swab samples (Suppl. material 2: fig. S3). In the honey samples, the highest proportion of reads assigned were to large subunit ribosomal ribonucleic acid (LSU rRNA). The proportion of all raw assigned reads that were identified as viruses ranged from 0.005% in the top box bees and honey samples, 0.016% in the hive entrance swabs, and 0.056% in the internal pollen samples.

After quality control and filtering steps, 67 viruses from a combined total of 17 million reads assigned to contigs were identified (Suppl. material 2: table S4). Total number of reads assigned to contigs was highest in the internal pollen samples, followed by the hive entrance swabs, top box bees, and the honey samples. Average contig lengths were highest in the internal pollen samples (3865.82 ± 3226.87 SD), followed by the top box bee samples (3544.36 ± 2925.95) and the hive entrance swabs (2833.81 ± 2743.46), and lowest in the honey samples (1298.83 ± 917.29 SD).

Honey samples had the highest observed number of viruses across all sample types, with an average number of 23 viruses per replicate (Fig. 7).

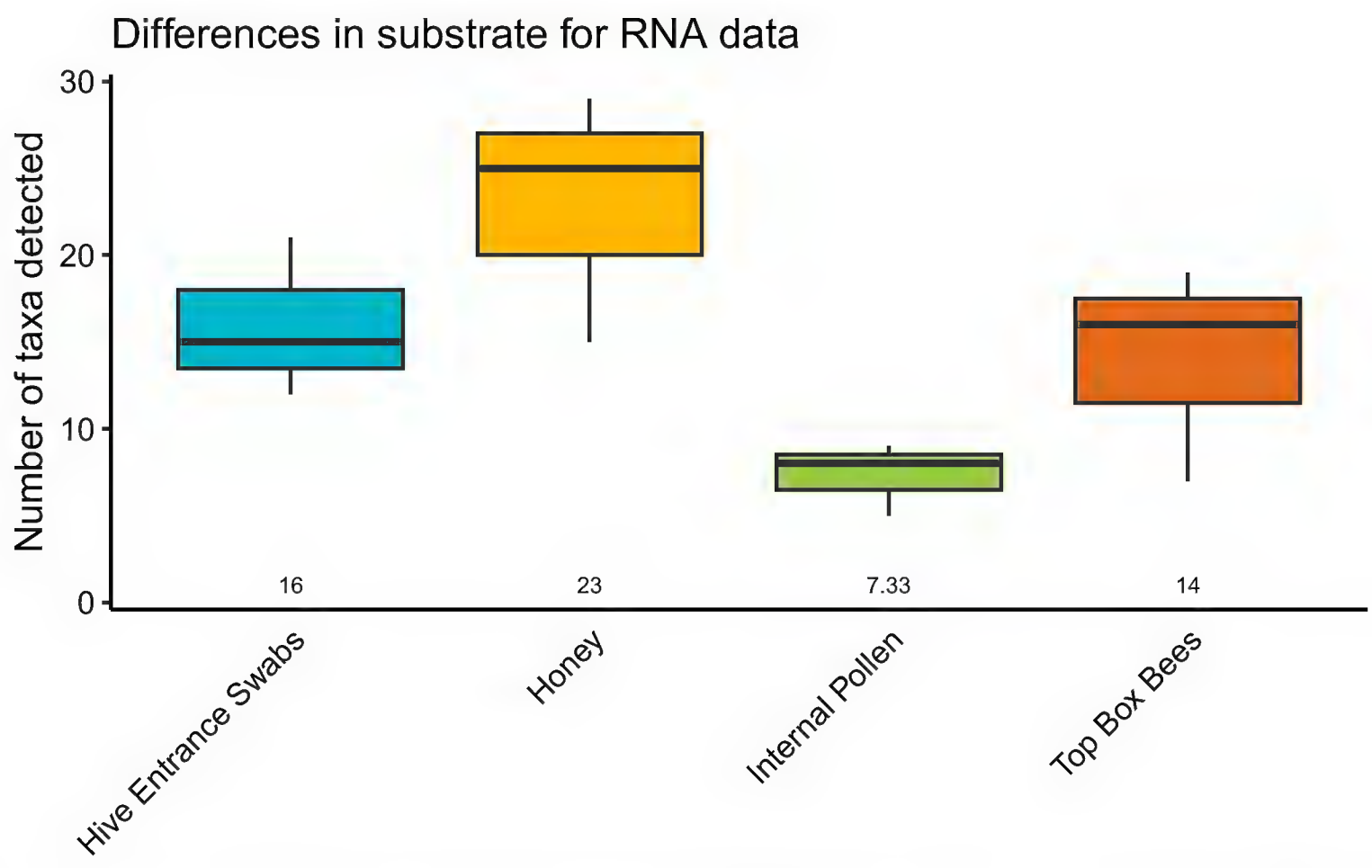

Figure 7. Number of taxa detected across each sample type for RNA viruses. Mean values for each substrate are shown below each boxplot.

This was followed by hive entrance swab and top box bee samples, with averages of 16and 14 viruses observed, respectively. Internal pollen samples contained the lowest observed number of viruses of any sample type with an average of 7.73.

Assigning detected viruses to expected host species, all four sample types were able to detect fungal and plant viruses, albeit at different rates (Fig. 8). Honey samples comprised the highest number of plant viruses with 23 plant viruses detected, followed by hive entrance swabs and top box bees (10 viruses each), and internal pollen samples (seven viruses). Fungal viruses were most abundant in swab samples (10 viruses), with two fungal viruses each in the bee and honey samples, and one fungal virus in the pollen samples. Honeybee-specific viruses were most abundant in the honeybee samples (seven viruses), followed by honey and hive entrance swab samples (six and three viruses, respectively). Other insect viruses were also detected in these sample types. No insect viruses were present in the internal pollen samples.

Five of the top 10 viruses detected across all sample types were associated with the plant genus *Camellia* (Fig. 9). These were all detected in each sample type, with the proportion across samples much higher in the internal pollen samples than the other sample types. Hive entrance swabs, honey, and top box bees showed positive detections of the remaining five most abundant viruses; these included two fungi-associated viruses (*Leptosphaeria biglobosa* betaflexivirus 3 and *Botryosphaeria dothidea* narnavirus 1), two viruses associated with clovers (white clover mosaic virus and clover yellow mosaic virus), and the honeybee virus black queen cell virus (*Triatovirus nigereginacellulae*). However, of these five viruses only one of the fungi viruses was detected in the internal pollen samples, resulting in only six of the top 10 viruses being detected in the internal pollen samples.

Shannon diversity estimates indicate that honey samples comprise the highest virus alpha diversity measures, followed by bee samples (Suppl. material 2: fig. S4). All hive entrance swab and pollen samples comprise virus alpha diversity values below 1.

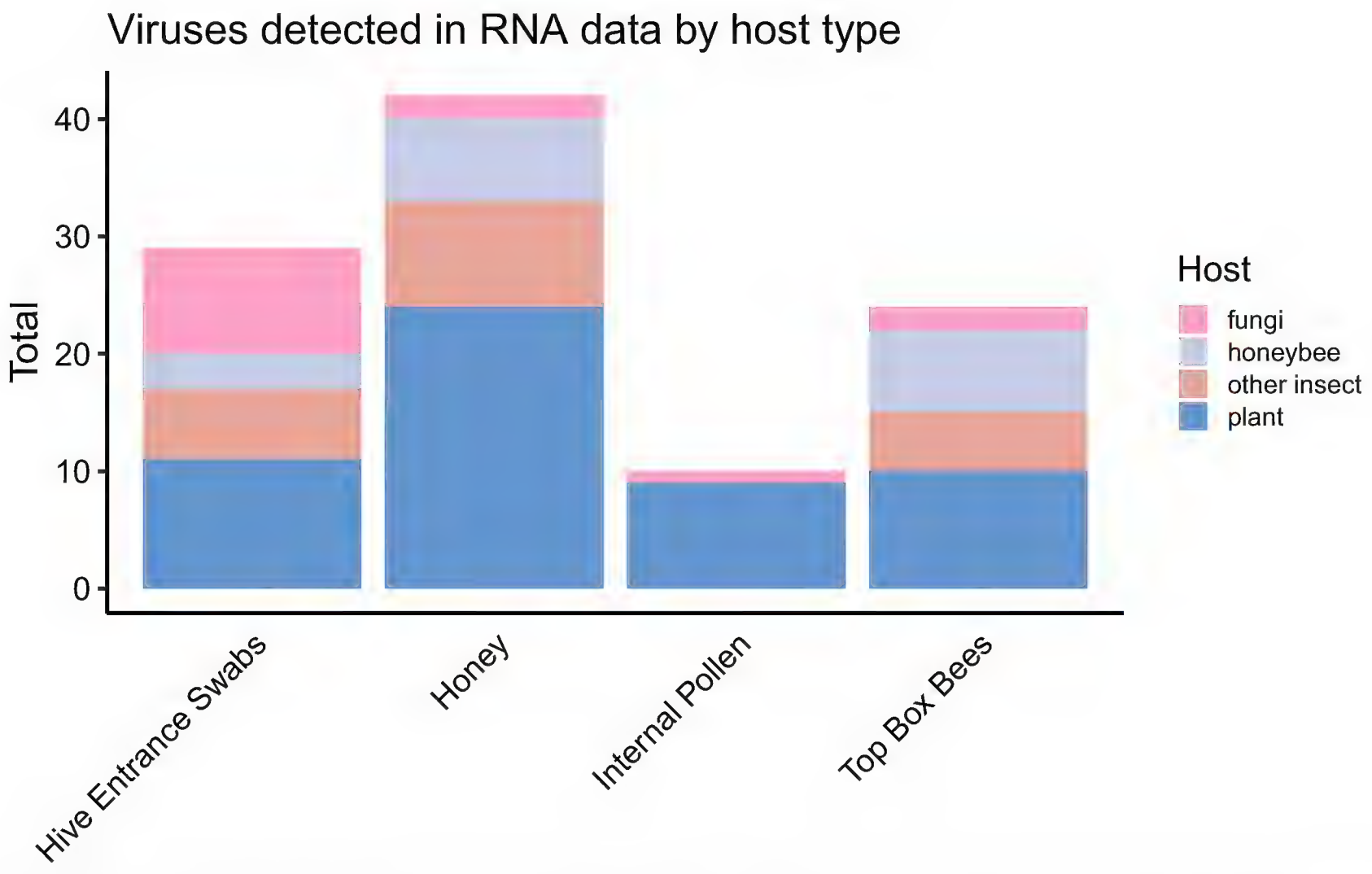

Figure 8. Number of viruses detected in different host categories across the four sample types sequenced.

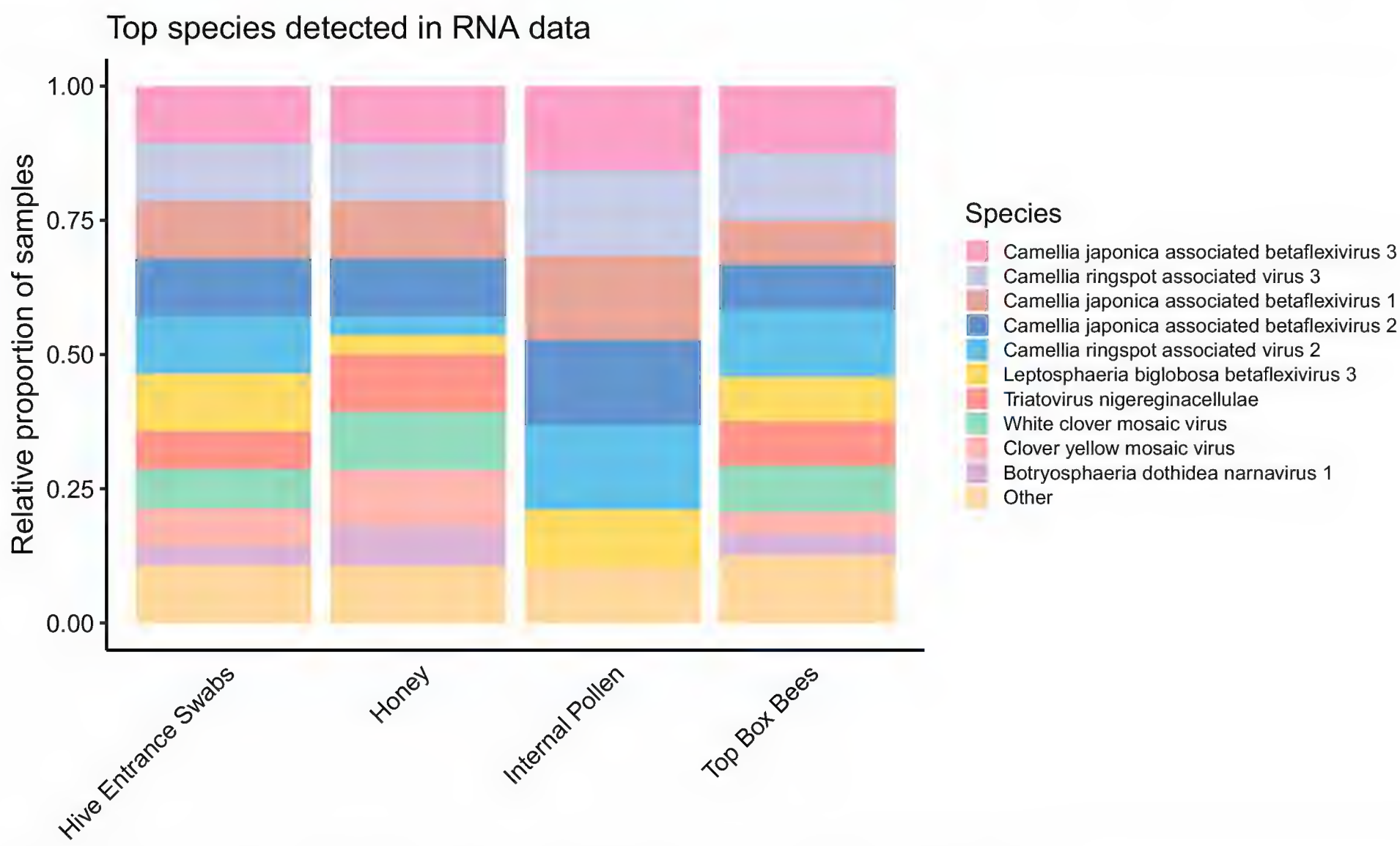

Figure 9. Top 10 viruses detected across all sample types and their detection proportions.

Discussion

Our study represents the most comprehensive comparison of hive eNA sample types to date. Expanding on the work of Boardman et al. (2024) and Čukanová et al. (2023) we examined all major taxonomic groups present in hives (bacteria,

animal, fungi, plants, viruses) to identify the most suitable sample types for honeybee pest and pathogen detection, vegetation monitoring and plant pathogen surveillance. We also draw on commonly utilised hive sample types (honey, pollen, bees) to provide a robust comparison of a non-invasive swabbing method to samples frequently used in the literature. Our results demonstrate that collecting swab samples from honeybee hive entrances is an effective method for detecting broad taxonomic assemblages of bacteria, animal, fungi, plants, and viruses. Further, we showed that the most common taxa and overall taxonomic composition are highly similar across sample types, indicating that swabs can reveal more diversity and the taxa identified are consistent with more intrusive sampling collection methods. From these results, we recommend hive entrance swabbing as the preferred method of sample collection from honeybee hives for metabarcoding studies, and a suitable method for RNA collection where the goal is broad-scale monitoring of viral diversity in the landscape. This work provides an improved honeybee hive sampling protocol that can be implemented for a range of biosecurity surveillance activities, including monitoring for honeybee related pests and pathogens, as well as plant pathogens and weeds.

There are clear benefits of hive entrance swabbing, being a simple method for non-specialists and enabling non-invasive year-round sample collection without impacting hives. The comparable and often better recovery of eNA from hive entrance swabs over more direct sample types was somewhat surprising. This effectiveness likely stems from the accumulation and persistence of eNA across the hive entrance and base. Hive debris typically accumulates at the base of the hive as a result of routine colony activities such as brood rearing and food storage. Additionally, high bee traffic through the hive entrance facilitates pollen, microbial spores and other biological material being deposited from returning foragers and passive air flow can also draw eNA from the surrounding area into hives. Other studies have also recognised the benefits of hive entrance swabbing for eDNA detection. Boardman et al. (2024) ranked hive entrance swabbing as one of their easy to collect sample methods that is ideal for large-scale monitoring, echoing our experience and findings presented here. Their comparison of swab and spray/wash methods from various hive environment sources found similar overall detection of arthropod and microbial taxa, although marginally improved species richness with spray/wash surface aggregation methods. Hive entrance swabbing was also recently demonstrated as an effective tool for targeted eDNA detection of honeybee pests and pathogens (Mackay et al. 2025; Roberts et al. 2025). While hive swabbing offers clear logistical advantages, it may not fully replace direct sampling methods for specific monitoring objectives, such as quantifying pathogen load or investigating hive foraging preferences. Additionally, factors such as hive design, climate, and seasonal variation are likely to influence eNA accumulation and detection sensitivity and warrant further investigation.

There were notable differences in the performance of the more common hive sample types (bees, pollen and honey) across taxonomic groups. Honey samples consistently showed high taxonomic diversity, suggesting they are effective for broad surveillance, similar to hive entrance swabs. The high sugar content of honey helps preserve eNA and has been used for detecting plant and microbial DNA even after several decades (Cirtwill and Wirta 2025). However, the honey samples also had low total read counts across taxa

groups, likely due to PCR inhibitors and lower eNA concentrations typical of this matrix (Soares et al. 2023). The relatively poor taxonomic recovery from the pollen samples, particularly for plants (ITS2), was unexpected. This may reflect the dynamic foraging behaviour of bees, which varies with the colony's nutritional needs (Coffey and Breen 1997; De Vere et al. 2017). Pollen and honey are also consumed by hives at different rates, with pollen typically used more quickly than honey stores (Anderson et al. 2014; Eyer et al. 2016; Carroll et al. 2017; Roessink and Van Der Steen 2021). As a result, honey likely provides a cumulative record of the local environment, while pollen samples give a contemporary view of recent flowering activity. Bee samples also showed lower species diversity, despite foraging bees being the primary mechanism for eNA entering the hive. Individual bees carry only a fraction of the total eNA, mostly their food sources and associated microbes, while surface-bound eNA is likely more transient. Bee samples also represent a narrower time window compared to honey and hive debris, which accumulate eNA over longer periods. These differences highlight how each sample type captures a distinct temporal and ecological snapshot of the hive environment, reinforcing the value of multi-sample approaches for comprehensive surveillance.

eRNA is increasingly used for distinguishing contemporary species presence and is essential for effective virus surveillance. Because RNA is more susceptible to degradation, direct sample types such as bees or pollen are generally expected to provide more reliable virus detection. However, our study found no optimal sample type for virus surveillance. Surprisingly, hive entrance swabs were comparable with honey and bee samples in capturing a broad snapshot of viral diversity, though they recovered fewer bee-specific viruses. Past investigations have also reported poor virus detection from hive entrance swabs (Čukanová et al. 2023). One possible explanation is that honeybees typically exit the hive for 'cleansing flights' to defecate, potentially reducing virus deposition on the hive base. Bee viruses may also degrade more rapidly in the environment than plant viruses, particularly those transmitted via pollen (Card et al. 2007). Honey samples provided the most comprehensive virus detection in terms of diversity, suggesting their suitability for holistic plant-pollinator virus surveillance. However, honey also yielded the lowest viral read counts and contig lengths, which may limit detection confidence. In contrast, bee and pollen samples produced higher read depths and contig lengths but detected fewer viruses overall. The lack of honeybee and other insect viruses detected in the pollen samples was also surprising. The recovery of plant viruses was comparable between sample types, suggesting there was no technical issue with the pollen RNA extractions. Multiple other studies also report success in detecting honeybee viruses using pollen samples (Pereira et al. 2019; Balkanska et al. 2023; Čukanová et al. 2023; Lee et al. 2023; Roberts et al. 2023; As and Eroglu 2025). It is likely our result is due to the small number of hives sampled and a general low prevalence of honeybee viruses in these hives at the time of sampling. These trade-offs between the number of viruses detected (quantity) and the depth and completeness of viral sequences (quality) are a key difference from DNA metabarcoding and emphasise that the most suitable sample type depends on the specific virus surveillance goal.

Optimising sampling effort is crucial for maximising taxonomic recovery in eNA studies, and our results provide specific guidance for honeybee hive

monitoring. Replication of swab samples in our study showed diminishing returns in the mean number of taxa and number of unique taxa by the third swab replicate, except for the fungal dataset. Therefore, we suggest that collecting two swab replicates per hive may be sufficient to capture the diversity of taxa from hive eDNA metabarcoding. For all the sample types and taxonomic groups we assessed via metabarcoding, additional sample and DNA extraction replicates led to increases in taxa recovery. However, the extent of replication necessary to capture the full breadth of taxa present in a sample or site will be dependent on both the study system being investigated and the taxonomic groups being amplified (Ficetola et al. 2015; Shirazi et al. 2021). Based on species accumulation curves, replication in our study was sufficient for the bacterial and animal communities, but further replication would be needed to recover additional diversity of the fungal and plant communities in our study hives. Greater sequencing depths will typically increase the detection of rare taxa in samples (Smith and Peay 2014; Doble et al. 2020; Shirazi et al. 2021), although our data suggests that this will still be more influenced by sample type. Using multiple primer pairs for target taxa can also circumvent possible biases and primer mismatches during PCR amplification (Freeland 2017; Xiong et al. 2022). Ultimately, the level of replication necessary for taxonomic recovery will be dependent on study questions, where aims of detecting rare taxa will require more replication than studies investigating broad-scale biodiversity. In honeybee hive surveillance, sampling from additional hives is likely to be more beneficial than collecting more than two swab replicates per hive and reduces antagonising a hive due to increased handling times.

Future research focusing on the ecology of eNA in relation to persistence within hives, variability across hives, and limits of detection is necessary to further improve the reliability and application of these biomonitoring methods. This is especially important in a biosecurity context where early detection of new incursions is critical and confidence around false negative and positive detections must be established. Further comparison of hive eNA detection with direct surveys would improve our understanding for how well hive taxonomic diversity provides a robust representation of the surrounding environment. Inclusion of eNA from external samples, such as from soil, water or air, over a range of distances from hives would help determine some of the practical limits of this approach. Hive-based detection is also inherently biased by bee foraging behaviour, which is influenced by floral availability, seasonality and landscape composition. Consideration of these factors will help refine when and where hive eNA surveillance is most effective. Despite these limitations in our understanding of hive eNA, the breadth of taxa detectable in bee hives is impressive and highlights a deeper complexity for how bees interact with their environment. One of the key strengths of eNA surveillance is the ability to conduct large-scale sampling efforts across diverse environments. Engaging with beekeepers to capture hive eNA across urban and agricultural landscapes is an exciting opportunity to support the health of hives and their foraging environments. Leveraging hive networks and simple swab methods offers a scalable, low-impact approach to landscape-level biosecurity and biodiversity monitoring, one that can be strengthened through collaboration with beekeepers and citizen scientists.

Conclusions

This study demonstrates that non-invasive swab sampling from hive entrances is an effective eNA method for detecting the broad diversity of taxa within honeybee hives. Hive entrance swabs were comparable with common hive eNA sampling methods, consistently recovering similar taxonomic profiles while also detecting additional taxa, particularly for bacterial and fungal groups. Sampling replication improved the number of taxa detected, with two swab replicates per hive recommended for future studies. Hive entrance swabbing also provides practical advantages in being simple, non-invasive, suitable for year-round sampling and can be conducted by non-specialists. These attributes, combined with broad taxonomic coverage, make hive entrance swabbing a powerful tool for large-scale biosecurity surveillance and biodiversity monitoring. Utilising established hive networks with this approach offers a highly scalable strategy for the detection of honeybee pests and pathogens, plant pathogens, and the surrounding vegetation across diverse landscapes.

Acknowledgements

The authors wish to thank Peter Jones, Biko Muita, Alicia Grealy and Harry Eyck for their assistance with sample collection and laboratory processing of samples. We are grateful to Florence Bravo and Amy Paten for their discussions regarding DNA and RNA extraction methods. This project was funded by the Environomics Future Science Platforms of the Commonwealth Scientific and Industrial Research Organization (CSIRO). We would like to acknowledge the contribution of Bioplatforms Australia in the generation of data used in this publication. Bioplatforms Australia is enabled by NCRIS.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Use of AI

No use of AI was reported.

Funding

No funding was reported.

Author contributions

Conceptualisation: EH, LM, FEV, BG, MH, JR. Data curation: EH, LM, JO. Formal analysis: EH, LM, JO. Funding acquisition: JR. Investigation: EH, LM, JR. Methodology: EH, LM, FEV, JO, JR. Software: LM, JO. Supervision: FEV, BG, MH, JR. Visualisation: EH, LM, JO. Writing – original draft: EH, LM, FEV, JO, JR. Writing – review and editing: EH, LM, FEV, JO, BG, MH, JR.

Author ORCIDs

Erin Hill https://orcid.org/0000-0002-7642-696X
Liz Milla https://orcid.org/0000-0002-6139-4336
Francisco Encinas-Viso https://orcid.org/0000-0003-0426-2342
James O'Dwyer https://orcid.org/0000-0003-3058-6295
Ben Gooden https://orcid.org/0000-0003-0575-9078
Mariana Hopper https://orcid.org/0000-0003-3685-473X
John Roberts https://orcid.org/0000-0001-9739-5595

Data availability

Raw fastq files, associated metadata and analysis pipelines can be accessed from the CSIRO Data Access Portal https://doi.org/10.25919/d2gq-vf73.

References

Abarenkov K, Nilsson RH, Larsson K-H, Taylor AFS, May TW, Frøslev TG, Pawlowska J, Lindahl B, Põldmaa K, Truong C, Vu D, Hosoya T, Niskanen T, Piirmann T, Ivanov F, Zirk A, Peterson M, Cheeke TE, Ishigami Y, Jansson AT, Jeppesen TS, Kristiansson E, Mikryukov V, Miller JT, Oono R, Ossandon FJ, Paupério J, Saar I, Schigel D, Suija A, Tedersoo L, Kõljalg U (2024) The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: Sequences, taxa and classifications reconsidered. Nucleic Acids Research 52(D1): D791–D797. https://doi.org/10.1093/nar/gkad1039

Anderson KE, Carroll MJ, Sheehan T, Mott BM, Maes P, Corby Harris V (2014) Hive stored pollen of honey bees: Many lines of evidence are consistent with pollen preservation, not nutrient conversion. Molecular Ecology 23(23): 5904–5917. https://doi.org/10.1111/mec.12966

Ankenbrand MJ, Keller A, Wolf M, Schultz J, Förster F (2015) ITS2 Database V: Twice as Much: Table 1. Molecular Biology and Evolution 32(11): 3030–3032. https://doi.org/10.1093/molbev/msv174

Apprill A, McNally S, Parsons R, Weber L (2015) Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquatic Microbial Ecology 75(2): 129–137. https://doi.org/10.3354/ame01753

As Y, Eroglu GB (2025) Molecular detection and phylogenetic analysis of five honey bee viruses from wet pollen in Türkiye. Biology Bulletin of the Russian Academy of Sciences 52(3): 89. https://doi.org/10.1134/S1062359024612898

Balkanska R, Shumkova R, Atsenova N, Salkova D, Dundarova H, Radoslavov G, Hristov P (2023) Molecular detection and phylogenetic analysis of deformed wing virus and sacbrood virus isolated from pollen. Veterinary Sciences 10(2): 140. https://doi.org/10.3390/vetsci10020140

Banerjee P, Stewart KA, Dey G, Antognazza CM, Sharma RK, Maity JP, Saha S, Doi H, De Vere N, Chan MWY, Lin P-Y, Chao H-C, Chen C-Y (2022) Environmental DNA analysis as an emerging non-destructive method for plant biodiversity monitoring: a review. AoB PLANTS 14: plac031. https://doi.org/10.1093/aobpla/plac031

Batchelor KL, Bell KL, Campos M, Webber BL (2023) Can honey bees be used to detect rare plants? Taking an EDNA approach to find the last plants in a weed eradication program. Environmental DNA 5(6): 1516–1526. https://doi.org/10.1002/edn3.471

Biová J, Charrière J-D, Dostálková S, Škrabišová M, Petřivalský M, Bzdil J, Danihlík J (2021) Melissococcus plutonius Can Be Effectively and Economically Detected Using

Hive Debris and Conventional PCR. Insects 12(2): 150. https://doi.org/10.3390/insects12020150

Boardman L, Marcelino JAP, Valentin RE, Boncristiani H, Standley JM, Ellis JD (2024) Novel EDNA approaches to monitor Western honey bee (*Apis mellifera* L.) microbial and arthropod communities. Environmental DNA 6(1): e419. https://doi.org/10.1002/edn3.419

Buchfink B, Reuter K, Drost H-G (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. Nature Methods 18(4): 366–368. https://doi.org/10.1038/s41592-021-01101-x

Bushnell B (2014) BBTools software packag. e.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: Architecture and applications. BMC Bioinformatics 10(1): 421. https://doi.org/10.1186/1471-2105-10-421

Card SD, Pearson MN, Clover GRG (2007) Plant pathogens transmitted by pollen. Australasian Plant Pathology 36(5): 455. https://doi.org/10.1071/AP07050

Carroll MJ, Brown N, Goodall C, Downs AM, Sheenan TH, Anderson KE (2017) Honey bees preferentially consume freshly-stored pollen. PLoS ONE 12: e0175933. https://doi.org/10.1371/journal.pone.0175933

Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo K, Li Y, Li X, Jia X, Lin Y, Leon C (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE 5: e8613. https://doi.org/10.1371/journal.pone.0008613

Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34(17): i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cirtwill AR, Wirta H (2025) DNA in honey could describe the changes in flower visits and microbe encounters of honey bees over decades. Scientific Reports 15(1): 8807. https://doi.org/10.1038/s41598-025-93315-8

Coffey MF, Breen J (1997) Seasonal variation in pollen and nectar sources of honey bees in Ireland. Journal of Apicultural Research 36(2): 63–76. https://doi.org/10.1080/00218839.1997.11100932

Čukanová E, Prodělalová J, Palíková M, Kováčová K, Linhart P, Papežíková I (2023) Can the examination of different types of hive samples be a non-invasive method for detection and quantification of viruses in honey bee (*Apis mellifera* L.) colonies? Journal of Veterinary Research (Pulawy) 67(3): 323–331. https://doi.org/10.2478/jvetres-2023-0046

Danner N, Keller A, Härtel S, Steffan-Dewenter I (2017) Honey bee foraging ecology: Season but not landscape diversity shapes the amount and diversity of collected pollen. PLoS ONE 12: e0183716. https://doi.org/10.1371/journal.pone.0183716

De Vere N, Jones LE, Gilmore T, Moscrop J, Lowe A, Smith D, Hegarty MJ, Creer S, Ford CR (2017) Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. Scientific Reports 7(1): 42838. https://doi.org/10.1038/srep42838

Doble CJ, Hipperson H, Salzburger W, Horsburgh GJ, Mwita C, Murrell DJ, Day JJ (2020) Testing the performance of environmental DNA metabarcoding for surveying highly diverse tropical fish communities: A case study from Lake Tanganyika. Environmental DNA 2(1): 24–41. https://doi.org/10.1002/edn3.43

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26(19): 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Edgar RC (2016) SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. https://doi.org/10.1101/074161

Eyer M, Neumann P, Dietemann V (2016) A look into the cell: Honey storage in honey bees, *Apis mellifera*. PLoS ONE 11: e0161059. https://doi.org/10.1371/journal.pone.0161059

Fetters AM, Ashman T (2023) The pollen virome: A review of pollen associated viruses and consequences for plants and their interactions with pollinators. American Journal of Botany 110(6): e16144. https://doi.org/10.1002/ajb2.16144

Ficetola GF, Pansu J, Bonin A, Coissac E, Giguet Covex C, De Barba M, Gielly L, Lopes CM, Boyer F, Pompanon F, Rayé G, Taberlet P (2015) Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular Ecology Resources 15(3): 543–556. https://doi.org/10.1111/1755-0998.12338

Fox J, Weisberg S (2018) An R companion to applied regression. Sage publications. https://doi.org/10.32614/CRAN.package.carData

Freeland JR (2017) The importance of molecular markers and primer design when characterizing biodiversity from environmental DNA. Genome 60: 358–374. https://doi.org/10.1139/gen-2016-0100

Gardes M, Bruns TD (1993) ITS primers with enhanced specificity for basidiomycetes application to the identification of mycorrhizae and rusts. Molecular Ecology 2(2): 113–118. https://doi.org/10.1111/j.1365-294X.1993.tb00005.x

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29(7): 644–652. https://doi.org/10.1038/nbt.1883

Hawkins J, De Vere N, Griffith A, Ford CR, Allainguillaume J, Hegarty MJ, Baillie L, Adams-Groom B (2015) Using DNA metabarcoding to identify the floral composition of honey: A new tool for investigating honey bee foraging preferences. PLoS ONE 10: e0134735. https://doi.org/10.1371/journal.pone.0134735

Hénaff E, Najjar D, Perez M, Flores R, Woebken C, Mason CE, Slavin K (2023) Holobiont Urbanism: Sampling urban beehives reveals cities' metagenomes. Environmental Microbiome 18(1): 23. https://doi.org/10.1186/s40793-023-00467-z

Kadlečková D, Tachezy R, Erban T, Deboutte W, Nunvář J, Saláková M, Matthijnssens J (2022) The Virome of Healthy Honey Bee Colonies: Ubiquitous Occurrence of Known and New Viruses in Bee Populations. mSystems 7: e00072-22. https://doi.org/10.1128/msystems.00072-22

Kassambara A (2019) rstatix: Pipe-friendly framework for basic statistical tests. CRAN: Contributed packages. https://doi.org/10.32614/CRAN.package.rstatix

Koetschan C, Förster F, Keller A, Schleicher T, Ruderisch B, Schwarz R, Müller T, Wolf M, Schultz J (2010) The ITS2 Database III—Sequences and structures for phylogeny. Nucleic Acids Research 38(suppl_1): D275–D279. https://doi.org/10.1093/nar/gkp966

Koetschan C, Hackl T, Müller T, Wolf M, Förster F, Schultz J (2012) ITS2 Database IV: Interactive taxon sampling for internal transcribed spacer 2 based phylogenies. Molecular Phylogenetics and Evolution 63(3): 585–588. https://doi.org/10.1016/j.ympev.2012.01.026

Lee E, Vansia R, Phelan J, Lofano A, Smith A, Wang A, Bilodeau GJ, Pernal SF, Guarna MM, Rott M, Griffiths JS (2023) Area Wide Monitoring of Plant and Honey Bee (*Apis mellifera*) Viruses in Blueberry (*Vaccinium corymbosum*) Agroecosystems Facilitated by Honey Bee Pollination. Viruses 15(5): 1209. https://doi.org/10.3390/v15051209

Leray M, Knowlton N, Machida RJ (2022) MIDORI2: A collection of quality controlled, preformatted, and regularly updated reference databases for taxonomic assignment

of eukaryotic mitochondrial sequences. Environmental DNA 4(4): 894–907. https://doi.org/10.1002/edn3.303

Mackay JF, Hewett RE, Smith NT, Waters TL, Scandrett JS (2025) The Foster method: Rapid and non-invasive detection of clinically significant American Foulbrood disease levels using eDNA sampling and a dual-target qPCR assay, with its potential for other hive pathogens. Journal of Apicultural Research 64(2): 693–700. https://doi.org/10.1080/00218839.2024.2306445

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal 17(1): 10–12. https://doi.org/10.14806/ej.17.1.200

McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, Zenger KR (2019) microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies. Environmental DNA 1(1): 14–25. https://doi.org/10.1002/edn3.11

McMurdie PJ, Holmes S (2013) phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8: e61217. https://doi.org/10.1371/journal.pone.0061217

Merget B, Koetschan C, Hackl T, Förster F, Dandekar T, Müller T, Schultz J, Wolf M (2012) The ITS2 Database. Journal of Visualized Experiments 3806(61): e3806. https://doi.org/10.3791/3806

Milla L, Sniderman K, Lines R, Mousavi Derazmahalleh M, Encinas Viso F (2021) Pollen DNA metabarcoding identifies regional provenance and high plant diversity in Australian honey. Ecology and Evolution 11(13): 8683–8698. https://doi.org/10.1002/ece3.7679

Milla L, Schmidt Lebuhn A, Bovill J, Encinas Viso F (2022) Monitoring of honey bee floral resources with pollen DNA metabarcoding as a complementary tool to vegetation surveys. Ecological Solutions and Evidence 3(1): e12120. https://doi.org/10.1002/2688-8319.12120

Minamoto T, Miya M, Sado T, Seino S, Doi H, Kondoh M, Nakamura K, Takahara T, Yamamoto S, Yamanaka H, Araki H, Iwasaki W, Kasai A, Masuda R, Uchii K (2021) An illustrated manual for environmental DNA research: Water sampling guidelines and experimental protocols. Environmental DNA 3(1): 8–13. https://doi.org/10.1002/edn3.121

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'hara R, Simpson GL, Solymos P (2019) vegan: Community Ecology Package. R package version 2.5-6.

Parada AE, Needham DM, Fuhrman JA (2016) Every base matters: Assessing small subunit RRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental Microbiology 18(5): 1403–1414. https://doi.org/10.1111/1462-2920.13023

Pawlowski J, Apothéloz-Perret-Gentil L, Mächler E, Altermatt F (2020) Environmental DNA applications for biomonitoring and bioassessment in aquatic ecosystems. https://doi.org/10.5167/UZH-187800

Pereira KDS, Meeus I, Smagghe G (2019) Honey bee-collected pollen is a potential source of Ascosphaera apis infection in managed bumble bees. Scientific Reports 9(1): 4241. https://doi.org/10.1038/s41598-019-40804-2

Piggott MP, Banks SC, Broadhurst BT, Fulton CJ, Lintermans M (2021) Comparison of traditional and environmental DNA survey methods for detecting rare and abundant freshwater fish. Aquatic Conservation 31(1): 173–184. https://doi.org/10.1002/aqc.3474

Prosser SWJ, Hebert PDN (2017) Rapid identification of the botanical and entomological sources of honey using DNA metabarcoding. Food Chemistry 214: 183–191. https://doi.org/10.1016/j.foodchem.2016.07.077

Ribani A, Utzeri VJ, Taurisano V, Fontanesi L (2020) Honey as a source of environmental dna for the detection and monitoring of honey bee pathogens and parasites. Veterinary Sciences 7(3): 113. https://doi.org/10.3390/vetsci7030113

Ribani A, Taurisano V, Utzeri VJ, Fontanesi L (2022) Honey environmental DNA can be used to detect and monitor honey bee pests: Development of methods useful to identify aethina tumida and galleria mellonella infestations. Veterinary Sciences 9(5): 213. https://doi.org/10.3390/vetsci9050213

Ribière C, Hegarty C, Stephenson H, Whelan P, O'Toole PW (2019) Gut and whole-body microbiota of the honey bee separate thriving and non-thriving hives. Microbial Ecology 78(1): 195–205. https://doi.org/10.1007/s00248-018-1287-9

Roberts JMK, Anderson DL, Durr PA (2018a) Metagenomic analysis of varroa-free australian honey bees (*Apis mellifera*) shows a diverse Picornavirales virome. The Journal of General Virology 99(6): 818–826. https://doi.org/10.1099/jgv.0.001073

Roberts JMK, Ireland KB, Tay WT, Paini D (2018b) Honey bee assisted surveillance for early plant virus detection. The Annals of Applied Biology 173(3): 285–293. https://doi.org/10.1111/aab.12461

Roberts JMK, Jooste AEC, Pretorius L-S, Geering ADW (2023) Surveillance for avocado sunblotch viroid utilizing the european honey bee (*Apis mellifera*). Phytopathology 113(3): 559–566. https://doi.org/10.1094/PHYTO-08-22-0295-R

Roberts JMK, Hall RJ, Shams F, Encinas Viso F, Bravo F, Soroka J, Milla L, Snape N, Martoni F, Walford A, Gleeson D, Trujillo González A (2025) Environmental DNA methods for detection of *Varroa destructor* in honey bee (*Apis mellifera*) hives. Environmental DNA 7(3): e70109. https://doi.org/10.1002/edn3.70109

Roessink I, Van Der Steen JJM (2021) Beebread consumption by honey bees is fast: Results of a six-week field study. Journal of Apicultural Research 60(5): 659–664. https://doi.org/10.1080/00218839.2021.1915612

Scriver M, Zaiko A, Pochon X, Von Ammon U (2023) Harnessing decay rates for coastal marine biosecurity applications: A review of environmental DNA and RNA fate. Environmental DNA 5(5): 960–972. https://doi.org/10.1002/edn3.405

Selig C, Wolf M, Muller T, Dandekar T, Schultz J (2007) The ITS2 Database II: Homology modelling RNA structure for molecular systematics. Nucleic Acids Research 36(Database): D377–D380. https://doi.org/10.1093/nar/gkm827

Shirazi S, Meyer RS, Shapiro B (2021) Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. Ecology and Evolution 11(22): 15766–15779. https://doi.org/10.1002/ece3.8239

Smith DP, Peay KG (2014) Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. PLoS ONE 9: e90234. https://doi.org/10.1371/journal.pone.0090234

Soares S, Rodrigues F, Delerue-Matos C (2023) Towards DNA-Based Methods Analysis for Honey: An Update. Molecules 28(5): 2106. https://doi.org/10.3390/molecules28052106

Tremblay ÉD, Duceppe M, Thurston GB, Gagnon M, Côté M, Bilodeau GJ (2019) High resolution biomonitoring of plant pathogens and plant species using metabarcoding of pollen pellet contents collected from a honey bee hive. Environmental DNA 1(2): 155–175. https://doi.org/10.1002/edn3.17

Trujillo-González A, Edmunds RC, Becker JA, Hutson KS (2019) Parasite detection in the ornamental fish trade using environmental DNA. Scientific Reports 9(1): 5173. https://doi.org/10.1038/s41598-019-41517-2

Utzeri VJ, Schiavo G, Ribani A, Tinarelli S, Bertolini F, Bovo S, Fontanesi L (2018) Entomological signatures in honey: An environmental DNA metabarcoding approach can

disclose information on plant-sucking insects in agricultural and forest landscapes. Scientific Reports 8(1): 9996. https://doi.org/10.1038/s41598-018-27933-w

Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp GH, Geniez P, Pont D, Argillier C, Baudoin J, Peroux T, Crivelli AJ, Olivier A, Acqueberge M, Le Brun M, Møller PR, Willerslev E, Dejean T (2016) Next generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular Ecology 25(4): 929–942. https://doi.org/10.1111/mec.13428

Vamos E, Elbrecht V, Leese F (2017) Short COI markers for freshwater macroinvertebrate metabarcoding. Metabarcoding and Metagenomics 1: e14625. https://doi.org/10.3897/mbmg.1.14625

Van Der Heyde M, Bunce M, Nevill P (2022) Key factors to consider in the use of environmental DNA metabarcoding to monitor terrestrial ecological restoration. The Science of the Total Environment 848: 157617. https://doi.org/10.1016/j.scitotenv.2022.157617

Veilleux HD, Misutka MD, Glover CN (2021) Environmental DNA and environmental RNA: Current and prospective applications for biological monitoring. The Science of the Total Environment 782: 146891. https://doi.org/10.1016/j.scitotenv.2021.146891

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (Eds) PCR Protocols. Academic Press, San Diego, 315–322. https://doi.org/10.1016/B978-0-12-372180-8.50042-1

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York [ISBN 978-3-319-24277-4].

Wirta H, Abrego N, Miller K, Roslin T, Vesterinen E (2021) DNA traces the origin of honey by identifying plants, bacteria and fungi. Scientific Reports 11(1): 4798. https://doi.org/10.1038/s41598-021-84174-0

Wood DE, Lu J, Langmead B (2019) Improved metagenomic analysis with Kraken 2. Genome Biology 20(1): 257. https://doi.org/10.1186/s13059-019-1891-0

Xiong F, Shu L, Zeng H, Gan X, He S, Peng Z (2022) Methodology for fish biodiversity monitoring with environmental DNA metabarcoding: The primers, databases and bioinformatic pipelines. Water Biology and Security 1(1): 100007. https://doi.org/10.1016/j.watbs.2022.100007

Yates MC, Derry AM, Cristescu ME (2021) Environmental RNA: A revolution in ecological resolution? Trends in Ecology & Evolution 36(7): 601–609. https://doi.org/10.1016/j.tree.2021.03.001

Supplementary material 1

PNA clamp development

Authors: Erin Hill, Liz Milla, Francisco Encinas-Viso, James O'Dwyer, Ben Gooden, Mariana Hopper, John Roberts

Data type: pdf

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.9.165436.suppl1

Supplementary material 2

Supporting tables and images

Authors: Erin Hill, Liz Milla, Francisco Encinas-Viso, James O'Dwyer, Ben Gooden, Mariana Hopper, John Roberts

Data type: pdf

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.9.165436.suppl2